OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 18:59:48 ; Search time 2686.26 Seconds
(without alignments)
185.904 Million cell updates/sec

Title: US-09-824-567-4
Perfect score: 37
Sequence: 1 gcgccggatcccatttttccttagcataacggaagtcc 37

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
	1	23.2	62.7	697	12	AZ433593	AZ433593 1M0219H04
	2	22.6	61.1	707	12	BH432458	BH432458 BOHAL85TR
c	3	22	59.5	409	10	BI476980	BI476980 daa87c05.
c	4	22	59.5	419	10	T06394	T06394 EST04283 Fe
	5	22	59.5	460	10	BG892973	BG892973 daa92b07.
	6	22	59.5	480	10	BJ059386	BJ059386 BJ059386
	7	22	59.5	573	10	BJ071059	BJ071059 BJ071059
	8	22	59.5	575	10	BJ069469	BJ069469 BJ069469
	9	22	59.5	616	10	BJ057427	BJ057427 BJ057427
	10	22	59.5	618	10	BJ096346	BJ096346 BJ096346
c	11	21.6	58.4	509	12	TA130F06Q	AL464265 T. brucei
c	12	21.2	57.3	708	12	AQ917850	AQ917850 RPCI-23-2
	13	21.2	57.3	741	10	BG432622	BG432622 602500787
	14	21	56.8	236	9	AV336182	AV336182 AV336182
c	15	21	56.8	440	12	AQ517828	AQ517828 HS_5133_B
c	16	21	56.8	466	12	AQ267958	AQ267958 RPCI11-72
c	17	21	56.8	556	12	BH063042	BH063042 RPCI-24-3
c	18	21	56.8	596	9	AU170167	AU170167 AU170167
	19	21	56.8	658	12	BH513817	BH513817 BOHLB55TR
c	20	20.8	56.2	477	9	BE022332	BE022332 sm73e08.y
c	21	20.8	56.2	481	9	BE021475	BE021475 sm59a03.y
	22	20.8	56.2	588	12	AZ863407	AZ863407 2M0171G02
c	23	20.8	56.2	788	10	BI907229	BI907229 603065275
	24	20.8	56.2	861	12	AZ670212	AZ670212 ENTKF96TR
	25	20.8	56.2	863	12	AZ679102	AZ679102 ENTIC54TR
c	26	20.8	56.2	883	12	AZ693290	AZ693290 ENTKB64TF
c	27	20.6	55.7	570	12	BH203311	BH203311 Sm1-43M11
c	28	20.6	55.7	578	12	AQ582878	AQ582878 RPCI-11-4
	29	20.6	55.7	638	12	AG140569	AG140569 Pan trogl
	30	20.6	55.7	673	9	AW689813	AW689813 NF024G03S
c	31	20.6	55.7	679	10	BJ071072	BJ071072 BJ071072
c	32	20.6	55.7	877	12	CNS03PNB	AL254864 Tetraodon
	33	20.6	55.7	1214	12	AG162839	AG162839 Pan trogl
c	34	20.4	55.1	279	10	BG143382	BG143382 mab57d01.
c	35	20.2	54.6	396	10	T96323	T96323 ye09d11.s1
c	36	20.2	54.6	403	10	T72691	T72691 yd19b02.r1
	37	20.2	54.6	667	12	BH041474	BH041474 RPCI-24-3
c	38	20.2	54.6	889	12	CNS0552A	AL321499 Tetraodon
c	39	20	54.1	162	10	R47338	R47338 He910-r Adu
	40	20	54.1	189	10	BG217964	BG217964 RST37688
c	41	20	54.1	224	10	BE996947	BE996947 NXCI_102_
c	42	20	54.1	417	9	AW801614	AW801614 IL5-UM006
	43	20	54.1	492	9	BE221089	BE221089 hv72c01.x
	44	20	54.1	500	12	AQ849056	AQ849056 LMAJFV1_1
	45	20	54.1	527	12	AQ474484	AQ474484 CITBI-E1-

ALIGNMENTS

RESULT 1
AZ433593
LOCUS AZ433593 697 bp DNA linear GSS 03-OCT-2000
DEFINITION 1M0219H04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0219H04 R, DNA sequence.
ACCESSION AZ433593
VERSION AZ433593.1 GI:10557606
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 697)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0219 row: H column: 04
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 697.
FEATURES Location/Qualifiers
source 1. .697
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0219H04"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 192 a 189 c 104 g 212 t
ORIGIN

Query Match 62.7%; Score 23.2; DB 12; Length 697;
Best Local Similarity 77.8%; Pred. No. 30;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy      1 gcgccggatcccattttccttagcataacggaaagtc 36
          | |||| |||| | || |||| || ||| |||||||
Db     16 GTGCCGAATCCTACTTCCCTTTGCCTAATGGAAGTC 51
```

RESULT 2
BH432458
LOCUS BH432458 707 bp DNA linear GSS 12-DEC-2001
DEFINITION BOHAL85TR BOHA Brassica oleracea genomic clone BOHAL85, DNA sequence.
ACCESSION BH432458
VERSION BH432458.1 GI:17618179
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 707)
AUTHORS Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: BOHAL85TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
FEATURES Location/Qualifiers
source 1. .707
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOHAL85"
/clone_lib="BOHA"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"

BASE COUNT 245 a 139 c 131 g 192 t
ORIGIN

Query Match 61.1%; Score 22.6; DB 12; Length 707;
Best Local Similarity 86.2%; Pred. No. 54;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      8 atcccatttttccttagcataacggaagtc 36
          || |||||||||| |||| |||| |||||
Db    215 ATACCATTTTCCTAAGCAAAACGTAAGTC 243
```

RESULT 3
BI476980/c
LOCUS BI476980 409 bp mRNA linear EST 27-AUG-2001
DEFINITION daa87c05.y4 Wellcome CRC pRN3 St13 17 egg animal cap Xenopus laevis cDNA clone IMAGE:4084232 5', mRNA sequence.
ACCESSION BI476980
VERSION BI476980.1 GI:15310396
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 409)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Xenopus EST project, 1999
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: daa87c05.x3
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by N. Garrett, E. ellefroid, and A.M. Zorn (Wellcome/CRC Institute). DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
High quality sequence stop: 383.
FEATURES Location/Qualifiers
source 1. .409
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4084232"
/clone_lib="Wellcome CRC pRN3 St13 17 egg animal cap"
/tissue_type="egg, subtracted by stage 13-17 animal cap"
/lab_host="DH10B (phage-resistant)"
/note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging according to Nieuwkoop and Faber. Library is subtracted and was constructed by N. Garrett, E Bellefroid, and A.M. Zorn, (Wellcome/CRC Institute). "

BASE COUNT 94 a 94 c 108 g 113 t
ORIGIN

Query Match 59.5%; Score 22; DB 10; Length 409;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy      4 ccggatcccatttttccttagcataacggaa 33
          |||||||||||| |||  |||| || ||||
Db    278 CCGGATCCCATTCTCCACAGCAGAATGGAA 249
```

RESULT 4
T06394/c
LOCUS T06394 419 bp mRNA linear EST 30-JUN-1993

```
DEFINITION  EST04283 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
            clone HFBDS84 similar to EST containing L1 repeat, mRNA sequence.
ACCESSION   T06394
VERSION     T06394.1  GI:317543
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
  TITLE     3,400 expressed sequence tags identify diversity of transcripts
            from human brain
  JOURNAL   Nature Genet. 4, 256-267 (1993)
  MEDLINE   93364420
COMMENT     Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):83059"
                     /db_xref="taxon:9606"
                     /clone="HFBDS84"
                     /clone_lib="Fetal brain, Stratagene (cat#936206)"
                     /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
                     oligo-dT + random primed cDNA synthesis; lambdaZAP-II
                     vector, 1.0kb average inser size."
BASE COUNT       90 a     86 c     78 g    156 t      9 others
ORIGIN
```

```
  Query Match           59.5%;  Score 22;  DB 10;  Length 419;
  Best Local Similarity  78.1%;  Pred. No. 87;
  Matches   25;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       3 gccggatcccatttccttagcataacggaag 34
           || ||| ||||||  |||||||| |||| | |
Db     384 GCTGGAGCCCATTNNCCTTAGCAAAACGAATG 353
```

```
RESULT   5
BG892973
LOCUS       BG892973                460 bp    mRNA    linear   EST 04-JUN-2001
DEFINITION  daa92b07.y1 Wellcome CRC pRN3 St13 17 egg animal cap Xenopus laevis
            cDNA clone IMAGE:4084573 5', mRNA sequence.
ACCESSION   BG892973
VERSION     BG892973.1  GI:14287583
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by N. Garrett, E. ellefroid, and A.M. Zorn
            (Wellcome/CRC Institute). DNA Sequencing by: Washington University
            Genome Sequencing Center
            Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            High quality sequence stop: 413.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4084573"
                     /clone_lib="Wellcome CRC pRN3 St13 17 egg animal cap"
                     /tissue_type="egg, subtracted by stage 13-17 animal cap"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
                     were oligo-dT primed and directionally cloned. Staging
                     according to  Nieuwkoop and Faber. Library is subtracted
                     and was constructed by N. Garrett, E Bellefroid, and A.M.
                     Zorn, (Wellcome/CRC Institute). "
BASE COUNT      106 a    110 c     99 g    145 t
ORIGIN
```

```
  Query Match           59.5%;  Score 22;  DB 10;  Length 460;
  Best Local Similarity  83.3%;  Pred. No. 89;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 ccggatcccatttccttagcataacggaa 33
           |||||||||||| |||  |||| || ||||
Db      29 CCGGATCCCATTCTCCACAGCAGAATGGAA 58
```

```
RESULT   6
BJ059386
LOCUS       BJ059386                480 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ059386 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL062h03 5', mRNA sequence.
ACCESSION   BJ059386
VERSION     BJ059386.1  GI:17492726
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL062h03"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
BASE COUNT      110 a    118 c    105 g    147 t
ORIGIN
```

```
  Query Match           59.5%;  Score 22;  DB 10;  Length 480;
  Best Local Similarity  83.3%;  Pred. No. 89;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;
```

```
Qy      4 ccggatcccatttttccttagcataacggaa 33
          ||||||||||||| |||  |||| || ||||
Db     77 CCGGATCCCATTCTCCACAGCAGAATGGAA 106
```

```
RESULT   7
BJ071059
LOCUS       BJ071059                  573 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ071059 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL092n13 5', mRNA sequence.
ACCESSION   BJ071059
VERSION     BJ071059.1  GI:17501248
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 573)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .573
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL092n13"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
BASE COUNT      136 a    134 c    123 g    180 t
ORIGIN
```

```
  Query Match           59.5%;  Score 22;  DB 10;  Length 573;
  Best Local Similarity  83.3%;  Pred. No. 92;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 ccggatcccatttttccttagcataacggaa 33
          ||||||||||||| |||  |||| || ||||
Db     57 CCGGATCCCATTCTCCACAGCAGAATGGAA 86
```

```
RESULT   8
BJ069469
LOCUS       BJ069469                  575 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ069469 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL053a06 5', mRNA sequence.
ACCESSION   BJ069469
VERSION     BJ069469.1  GI:17497829
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 575)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL053a06"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
BASE COUNT      162 a    132 c    136 g    143 t      2 others
ORIGIN
```

```
  Query Match           59.5%;  Score 22;  DB 10;  Length 575;
  Best Local Similarity  83.3%;  Pred. No. 92;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 ccggatcccatttttccttagcataacggaa 33
          ||||||||||||| |||  |||| || ||||
Db    223 CCGGATCCCATTCTCCACAGCAGAATGGAA 252
```

```
RESULT   9
BJ057427
LOCUS       BJ057427                  616 bp    mRNA    linear   EST 10-DEC-2001
DEFINITION  BJ057427 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL104a11 5', mRNA sequence.
ACCESSION   BJ057427
VERSION     BJ057427.1  GI:17470221
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .616
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL104a11"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
BASE COUNT      188 a    150 c    161 g    115 t      2 others
ORIGIN
```

```
  Query Match           59.5%;  Score 22;  DB 10;  Length 616;
  Best Local Similarity  83.3%;  Pred. No. 93;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 ccggatcccatttttccttagcataacggaa 33
          ||||||||||||| |||  |||| || ||||
Db    399 CCGGATCCCATTCTCCACAGCAGAATGGAA 428
```

```
RESULT 10
BJ096346
LOCUS       BJ096346                618 bp    mRNA    linear   EST 12-DEC-2001
DEFINITION  BJ096346 NIBB Mochii normalized Xenopus early gastrula library
            Xenopus laevis cDNA clone XL153j23 5', mRNA sequence.
ACCESSION   BJ096346
VERSION     BJ096346.1  GI:17597224
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .618
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL153j23"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
BASE COUNT      181 a    150 c    150 g    137 t
ORIGIN
```

```
  Query Match           59.5%;  Score 22;  DB 10;  Length 618;
  Best Local Similarity  83.3%;  Pred. No. 93;
  Matches   25;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 ccggatcccatttccttagcataacggaa 33
           ||||||||||||| |||  |||| || ||||
Db     316 CCGGATCCCATTCTCCACAGCAGAATGGAA 345
```

```
RESULT 11
TA130F06Q/c
LOCUS       TA130F06Q               509 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 130f06, reverse sequence,
            genomic survey sequence.
ACCESSION   AL464265
VERSION     AL464265.1  GI:11834528
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="130f06"
BASE COUNT       97 a    130 c    122 g    160 t
ORIGIN
```

```
  Query Match           58.4%;  Score 21.6;  DB 12;  Length 509;
  Best Local Similarity  75.0%;  Pred. No. 1.3e+02;
  Matches   27;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 gcgccggatcccatttccttagcataacggaagtc 36
           | |||| || | ||| | |||| ||||||| ||| |
Db     460 GAGCCGAATACGATTATACTTATCATAACGAAAGCC 425
```

```
RESULT 12
AQ917850/c
LOCUS       AQ917850                708 bp    DNA     linear   GSS 21-DEC-1999
DEFINITION  RPCI-23-285D14.TJ RPCI-23 Mus musculus genomic clone RPCI-23-285D14
            , DNA sequence.
ACCESSION   AQ917850
VERSION     AQ917850.1  GI:6606852
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 708)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-285D14.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 285  row: D  column: 14
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .708
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-285D14"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
```

with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

BASE COUNT 152 a 162 c 167 g 227 t
ORIGIN

Query Match 57.3%; Score 21.2; DB 12; Length 708;
Best Local Similarity 76.5%; Pred. No. 2e+02;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 2 cgccggatcccattttccttagcataacggaagt 35
Db 613 CCCCGAATCCCATGTGCCTGAGCATCACAGAAAT 580

RESULT 13
BG432622
LOCUS BG432622 741 bp mRNA linear EST 14-MAR-2001
DEFINITION 602500787F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4614356 5', mRNA sequence.
ACCESSION BG432622
VERSION BG432622.1 GI:13339128
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 741)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLCM1365 row: n column: 21
High quality sequence stop: 558.
FEATURES Location/Qualifiers
source 1. .741
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4614356"
/clone_lib="NIH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT 211 a 129 c 152 g 248 t 1 others
ORIGIN

Query Match 57.3%; Score 21.2; DB 10; Length 741;
Best Local Similarity 76.5%; Pred. No. 2e+02;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 4 ccggatcccattttccttagcataacggaagtcc 37
Db 640 CTGGTAACTATTTTCCTTAATATAACGGAATTCC 673

RESULT 14
AV336182
LOCUS AV336182 236 bp mRNA linear EST 11-NOV-1999
DEFINITION AV336182 RIKEN full-length enriched, adult male medulla oblongata Mus musculus cDNA clone 6330583D01 3', mRNA sequence.
ACCESSION AV336182
VERSION AV336182.1 GI:6376234
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 236)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al. 1999)
JOURNAL Unpublished (1999)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .236
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6330583D01"
/clone_lib="RIKEN full-length enriched, adult male medulla oblongata"
/sex="male"
/tissue_type="medulla oblongata"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization

to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT 60 a 47 c 31 g 98 t
ORIGIN

Query Match 56.8%; Score 21; DB 9; Length 236;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 gatcccattttccttagcata 27
||||||||||||||||||||
Db 180 GATCCCATTTTCCTTAGCATA 200

RESULT 15
AQ517828/c
LOCUS AQ517828 440 bp DNA linear GSS 05-MAY-1999
DEFINITION HS_5133_B2_D03_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=709 Col=6 Row=H, DNA sequence.
ACCESSION AQ517828
VERSION AQ517828.1 GI:4749976
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 440)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 709 row: H column: 6
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 440.
FEATURES Location/Qualifiers
source 1. .440
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=709 Col=6 Row=H"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 76 a 143 c 131 g 88 t 2 others
ORIGIN

Query Match 56.8%; Score 21; DB 12; Length 440;
Best Local Similarity 73.0%; Pred. No. 2.3e+02;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 gcgccggatcccatttttccttagcataacggaagtcc 37
||||| ||||| | ||| ||| ||||| | ||||
Db 75 GCGCCCGATCCGACCTTCTTTATTATAACAAATGTCC 39

Search completed: July 31, 2002, 18:59:52
Job time: 6634 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 19:32:59 ; Search time 84.08 Seconds
(without alignments)
108.093 Million cell updates/sec

Title: US-09-824-567-4
Perfect score: 37
Sequence: 1 gcgccggatcccattttccttagcataacggaagtcc 37

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	19.8	53.5	497	3	US-09-248-528-4	Sequence 4, Appli
	2	19.8	53.5	497	4	US-09-549-108-4	Sequence 4, Appli
	3	19.8	53.5	497	4	US-09-549-111-4	Sequence 4, Appli
	4	19.8	53.5	497	4	US-09-549-106-4	Sequence 4, Appli
	5	19.8	53.5	497	4	US-09-550-394-4	Sequence 4, Appli
	6	18.2	49.2	10803	3	US-09-080-044-1	Sequence 1, Appli
	7	17.8	48.1	30	4	US-09-522-666-13	Sequence 13, Appl
	8	17.8	48.1	2859	2	US-08-637-763B-7	Sequence 7, Appli
	9	17.8	48.1	2859	3	US-09-170-354-7	Sequence 7, Appli
c	10	17.6	47.6	3092	4	US-09-522-666-1	Sequence 1, Appli
	11	17.4	47.0	6464	1	US-08-321-478-2	Sequence 2, Appli
	12	17.4	47.0	6464	1	US-08-321-478-4	Sequence 4, Appli
	13	17.4	47.0	6464	1	US-08-321-478-6	Sequence 6, Appli
c	14	17.4	47.0	10754	2	US-08-966-958-1	Sequence 1, Appli
c	15	17.4	47.0	10754	2	US-09-215-817-1	Sequence 1, Appli
c	16	17.4	47.0	10754	4	US-09-342-353-1	Sequence 1, Appli
	17	17.4	47.0	11236	1	US-07-853-913-1	Sequence 1, Appli
	18	17.2	46.5	593	4	US-09-328-111-724	Sequence 724, App
c	19	17.2	46.5	607	3	US-08-894-483-6	Sequence 6, Appli
c	20	17.2	46.5	4895	4	US-09-426-568A-3	Sequence 3, Appli
c	21	17	45.9	573	4	US-09-385-982-451	Sequence 451, App
c	22	17	45.9	830	4	US-08-998-416-298	Sequence 298, App
c	23	17	45.9	1460	1	US-08-133-038A-1	Sequence 1, Appli
c	24	17	45.9	1460	1	US-08-161-988A-1	Sequence 1, Appli
c	25	17	45.9	2149	1	US-08-784-651-3	Sequence 3, Appli
	26	17	45.9	8752	4	US-08-976-259-3	Sequence 3, Appli
	27	17	45.9	72928	3	US-09-009-913-1	Sequence 1, Appli
	28	16.8	45.4	666	1	US-08-663-023-16	Sequence 16, Appl
	29	16.8	45.4	2362	1	US-08-265-087-1	Sequence 1, Appli
	30	16.8	45.4	2362	1	US-08-621-493-1	Sequence 1, Appli
	31	16.8	45.4	2362	2	US-08-965-688-1	Sequence 1, Appli
	32	16.8	45.4	2362	4	US-09-260-173-1	Sequence 1, Appli
c	33	16.8	45.4	2458	3	US-08-611-587-6	Sequence 6, Appli
	34	16.8	45.4	2994	1	US-08-204-329-2	Sequence 2, Appli
	35	16.8	45.4	2994	2	US-08-482-627-4	Sequence 4, Appli
	36	16.8	45.4	2994	3	US-08-801-092-3	Sequence 3, Appli
	37	16.8	45.4	2994	5	PCT-US94-10357-1	Sequence 1, Appli
	38	16.8	45.4	2995	2	US-08-959-638-7	Sequence 7, Appli
	39	16.8	45.4	2995	4	US-08-328-673A-7	Sequence 7, Appli
	40	16.8	45.4	3153	4	US-09-175-928-9	Sequence 9, Appli
	41	16.8	45.4	3232	1	US-08-038-760-1	Sequence 1, Appli
c	42	16.8	45.4	3232	1	US-08-038-760-2	Sequence 2, Appli
	43	16.8	45.4	3232	2	US-08-470-091-1	Sequence 1, Appli
c	44	16.8	45.4	3232	2	US-08-470-091-2	Sequence 2, Appli
c	45	16.8	45.4	3300	3	US-08-913-842-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT   1
US-09-248-528-4
; Sequence 4, Application US/09248528C
; Patent No. 6153415
; GENERAL INFORMATION:
;  APPLICANT: Oriel, Patrick J
;  APPLICANT: Padmakumar, Rugmini
;  APPLICANT: Kim, Sang H
;  TITLE OF INVENTION: Method for Producing Amide Compounds Using a  Nitrile
;  TITLE OF INVENTION: Hydratase from a Thermophilic Bacillus
;  FILE REFERENCE: MSU 4.1-401
;  CURRENT APPLICATION NUMBER: US/09/248,528C
;  CURRENT FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: 60/083,485
;  EARLIER FILING DATE: 1998-04-29
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 497
;   TYPE: DNA
;   ORGANISM: Bacillus smithii
;   FEATURE:
;   NAME/KEY: rRNA
;   LOCATION: (1)..(497)
;   OTHER INFORMATION: nucleotides 17 - 513 of rRNA sequence
;   OTHER INFORMATION: X60643/Genbank
;  PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: X60643/Genbank
;   DATABASE ENTRY DATE: 1997-04-03
US-09-248-528-4


  Query Match           53.5%;  Score 19.8;  DB 3;  Length 497;
  Best Local Similarity  77.4%;  Pred. No. 5.1;
  Matches   24;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 ccggatcccattttccttagcataacggaag 34
           ||||||   || |||||| |||| | |||||
Db     152 ccggataatatcttccttcgcatgaaggaag 182


RESULT   2
US-09-549-108-4
; Sequence 4, Application US/09549108
; Patent No. 6214603
; GENERAL INFORMATION:
;  APPLICANT:  Oriel, Patrick J
;  APPLICANT:  Padmakumar, Rugmini
;  APPLICANT:  Kim, Sang H
;  TITLE OF INVENTION:  Method for Producing Amide Compounds Using a  Nitrile
```

```
;  TITLE OF INVENTION:  Hydratase from a Thermophilic Bacillus
;  FILE REFERENCE:  MSU 4.1-486
;  CURRENT APPLICATION NUMBER: US/09/549,108
;  CURRENT FILING DATE:  2000-04-13
;  PRIOR APPLICATION NUMBER:  60/083,485
;  PRIOR FILING DATE:  1998-04-29
;  PRIOR APPLICATION NUMBER:  09/248,528
;  PRIOR FILING DATE:  1999-02-10
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 4
;    LENGTH: 497
;    TYPE: DNA
;    ORGANISM: Bacillus smithii
;    FEATURE:
;    NAME/KEY: rRNA
;    LOCATION: (1)..(497)
;    OTHER INFORMATION: nucleotides 17 - 513 of rRNA sequence
;    OTHER INFORMATION: X60643/Genbank
;  PUBLICATION INFORMATION:
;    DATABASE ACCESSION NUMBER: X60643/Genbank
;    DATABASE ENTRY DATE: 1997-04-03
US-09-549-108-4


  Query Match           53.5%;  Score 19.8;  DB 4;  Length 497;
  Best Local Similarity  77.4%;  Pred. No. 5.1;
  Matches   24;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 ccggatcccattttccttagcataacggaag 34
           ||||||   || |||||| |||| | |||||
Db     152 ccggataatatcttccttcgcatgaaggaag 182


RESULT   3
US-09-549-111-4
; Sequence 4, Application US/09549111
; Patent No. 6228633
; GENERAL INFORMATION:
;  APPLICANT:  Oriel, Patrick J
;  APPLICANT:  Padmakumar, Rugmini
;  APPLICANT:  Kim, Sang H
;  TITLE OF INVENTION:  Method for Producing Amide Compounds Using a  Nitrile
;  TITLE OF INVENTION:  Hydratase from a Thermophilic Bacillus
;  FILE REFERENCE:  MSU 4.1-489
;  CURRENT APPLICATION NUMBER: US/09/549,111
;  CURRENT FILING DATE:  2000-04-13
;  PRIOR APPLICATION NUMBER:  60/083,485
;  PRIOR FILING DATE:  1998-04-29
;  PRIOR APPLICATION NUMBER:  09/248,528
;  PRIOR FILING DATE:  1999-02-10
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 4
;    LENGTH: 497
;    TYPE: DNA
;    ORGANISM: Bacillus smithii
;    FEATURE:
;    NAME/KEY: rRNA
;    LOCATION: (1)..(497)
;    OTHER INFORMATION: nucleotides 17 - 513 of rRNA sequence
;    OTHER INFORMATION: X60643/Genbank
;  PUBLICATION INFORMATION:
;    DATABASE ACCESSION NUMBER: X60643/Genbank
;    DATABASE ENTRY DATE: 1997-04-03
US-09-549-111-4


  Query Match           53.5%;  Score 19.8;  DB 4;  Length 497;
  Best Local Similarity  77.4%;  Pred. No. 5.1;
  Matches   24;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 ccggatcccattttccttagcataacggaag 34
           ||||||   || |||||| |||| | |||||
Db     152 ccggataatatcttccttcgcatgaaggaag 182


RESULT   4
US-09-549-106-4
; Sequence 4, Application US/09549106
; Patent No. 6242242
; GENERAL INFORMATION:
;  APPLICANT:  Oriel, Patrick J
;  APPLICANT:  Padmakumar, Rugmini
;  APPLICANT:  Kim, Sang H
;  TITLE OF INVENTION:  Method for Producing Amide Compounds Using a  Nitrile
;  TITLE OF INVENTION:  Hydratase from a Thermophilic Bacillus
;  FILE REFERENCE:  MSU 4.1-487
;  CURRENT APPLICATION NUMBER: US/09/549,106
;  CURRENT FILING DATE:  2000-04-13
;  PRIOR APPLICATION NUMBER:  60/083,485
;  PRIOR FILING DATE:  1998-04-29
;  PRIOR APPLICATION NUMBER:  09/248,528
;  PRIOR FILING DATE:  1999-02-10
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 4
;    LENGTH: 497
;    TYPE: DNA
;    ORGANISM: Bacillus smithii
;    FEATURE:
;    NAME/KEY: rRNA
;    LOCATION: (1)..(497)
;    OTHER INFORMATION: nucleotides 17 - 513 of rRNA sequence
;    OTHER INFORMATION: X60643/Genbank
;  PUBLICATION INFORMATION:
;    DATABASE ACCESSION NUMBER: X60643/Genbank
;    DATABASE ENTRY DATE: 1997-04-03
US-09-549-106-4


  Query Match           53.5%;  Score 19.8;  DB 4;  Length 497;
  Best Local Similarity  77.4%;  Pred. No. 5.1;
  Matches   24;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       4 ccggatcccattttccttagcataacggaag 34
           ||||||   || |||||| |||| | |||||
Db     152 ccggataatatcttccttcgcatgaaggaag 182


RESULT   5
US-09-550-394-4
; Sequence 4, Application US/09550394
; Patent No. 6287828
; GENERAL INFORMATION:
;  APPLICANT:  Oriel, Patrick J
;  APPLICANT:  Padmakumar, Rugmini
;  APPLICANT:  Kim, Sang H
;  TITLE OF INVENTION:  Method for Producing Amide Compounds Using a  Nitrile
;  TITLE OF INVENTION:  Hydratase from a Thermophilic Bacillus
;  FILE REFERENCE:  MSU 4.1-488
;  CURRENT APPLICATION NUMBER: US/09/550,394
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER:  60/083,485
;  PRIOR FILING DATE:  1998-04-29
;  PRIOR APPLICATION NUMBER:  09/248,528
;  PRIOR FILING DATE:  1999-02-10
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 4
;    LENGTH: 497
;    TYPE: DNA
;    ORGANISM: Bacillus smithii
;    FEATURE:
;    NAME/KEY: rRNA
;    LOCATION: (1)..(497)
```

; OTHER INFORMATION: nucleotides 17 - 513 of rRNA sequence
; OTHER INFORMATION: X60643/Genbank
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: X60643/Genbank
; DATABASE ENTRY DATE: 1997-04-03
US-09-550-394-4

Query Match 53.5%; Score 19.8; DB 4; Length 497;
Best Local Similarity 77.4%; Pred. No. 5.1;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 4 ccggatcccattttccttagcataacggaag 34
|||||| || |||||| |||| | |||||
Db 152 ccggataatatcttccttcgcatgaaggaag 182

RESULT 6
US-09-080-044-1
; Sequence 1, Application US/09080044
; Patent No. 6074649
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, Jean-Christophe F.
; APPLICANT: BAUDU, Philippe G.
; APPLICANT: RIVIERE, Michel A.
; TITLE OF INVENTION: RECOMBINANT VACCINE CONTAINING FELINE HERPES VIRUS TYPE
; TITLE OF INVENTION: 1, PARTICULARLY FOR TREATING FELINE INFECTIOUS
; TITLE OF INVENTION: PERITONITIS
; FILE REFERENCE: AUDONNET
; CURRENT APPLICATION NUMBER: US/09/080,044
; CURRENT FILING DATE: 1998-05-15
; EARLIER APPLICATION NUMBER: PCT/FR96/01830
; EARLIER FILING DATE: 1996-11-19
; EARLIER APPLICATION NUMBER: 95/14450
; EARLIER FILING DATE: 1995-11-30
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 10803
; TYPE: DNA
; ORGANISM: Feline herpesvirus 1
US-09-080-044-1

Query Match 49.2%; Score 18.2; DB 3; Length 10803;
Best Local Similarity 74.2%; Pred. No. 53;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 3 gccggatcccattttccttagcataacggaa 33
|||| ||||||| || | ||| | | ||||
Db 310 gccgcatcccatctttcgaagctttatggaa 340

RESULT 7
US-09-522-666-13
; Sequence 13, Application US/09522666
; Patent No. 6333167
; GENERAL INFORMATION:
; APPLICANT: Shuey, David
; APPLICANT: Quinet, Elaine
; TITLE OF INVENTION: Methods and Reagents for Identifying Inhibitors of
; TITLE OF INVENTION: Proteolysis of Membrane-Associated Proteins
; FILE REFERENCE: 6-00
; CURRENT APPLICATION NUMBER: US/09/522,666
; CURRENT FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 30
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:oligonucleotide
US-09-522-666-13

Query Match 48.1%; Score 17.8; DB 4; Length 30;
Best Local Similarity 75.9%; Pred. No. 21;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 gcgccggatcccattttccttagcataac 29
|| |||||||| | ||| | ||||| ||
Db 1 gctccggatcctttcttcttcagcatcac 29

RESULT 8
US-08-637-763B-7
; Sequence 7, Application US/08637763B
; Patent No. 5849559
; GENERAL INFORMATION:
; APPLICANT: VAN DER WOUW, Monique J.A. et al
; TITLE OF INVENTION: ARABINOXYLAN DEGRADING ENZYME
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,763B
; FILING DATE: 25-AUG-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 4615-0066.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus niger var. tubigensis
; STRAIN: DS16813
; FEATURE:
; NAME/KEY: CAAT_signal
; LOCATION: 651..655
; FEATURE:
; NAME/KEY: TATA_signal
; LOCATION: 713..720
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 823..1818
; OTHER INFORMATION: /product= "arabinoxylan degrading
; OTHER INFORMATION: enzyme"
; OTHER INFORMATION: /gene= "axdA"
; OTHER INFORMATION: /standard_name= "arabinoxylan degrading enzyme"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 823..901
; FEATURE:

```
;       NAME/KEY:  mat_peptide
;       LOCATION:  901..1818
US-08-637-763B-7


  Query Match             48.1%;  Score 17.8;  DB 2;  Length 2859;
  Best Local Similarity   67.6%;  Pred. No. 59;
  Matches   25;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 gcgccggatcccatttttccttagcataacggaagtcc 37
           ||| |||||||||||| | ||   ||    |||| | |
Db    2551 GCGTCGGATCCCATTGTGCTCGACAGCCTGGAACTTC 2587


RESULT   9
US-09-170-354-7
; Sequence 7, Application US/09170354
; Patent No. 6066356
;  GENERAL INFORMATION:
;     APPLICANT:  VAN DER WOUW, Monique J.A. et al
;     TITLE OF INVENTION:  ARABINOXYLAN DEGRADING ENZYME
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Morrison & Foerster
;       STREET:  2000 Pennsylvania Avenue, NW
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20006-1812
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/170,354
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/637,763
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Murashige, Kate H.
;       REGISTRATION NUMBER:  29,959
;       REFERENCE/DOCKET NUMBER:  4615-0066.00
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 887-1500
;       TELEFAX:  (202) 887-0763
;       TELEX:  90-4030 MRSNFOERSWSH
;  INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2859 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Aspergillus niger var. tubigensis
;       STRAIN:  DS16813
;     FEATURE:
;       NAME/KEY:  CAAT_signal
;       LOCATION:  651..655
;     FEATURE:
;       NAME/KEY:  TATA_signal
;       LOCATION:  713..720
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  823..1818
;       OTHER INFORMATION:  /product= "arabinoxylan degrading
;       OTHER INFORMATION:  enzyme"
;       OTHER INFORMATION:  /gene= "axdA"
;       OTHER INFORMATION:  /standard_name= "arabinoxylan degrading enzyme"
;     FEATURE:
;       NAME/KEY:  sig_peptide
;       LOCATION:  823..901
;     FEATURE:
;       NAME/KEY:  mat_peptide
;       LOCATION:  901..1818
US-09-170-354-7


  Query Match             48.1%;  Score 17.8;  DB 3;  Length 2859;
  Best Local Similarity   67.6%;  Pred. No. 59;
  Matches   25;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 gcgccggatcccatttttccttagcataacggaagtcc 37
           ||| |||||||||||| | ||   ||    |||| | |
Db    2551 GCGTCGGATCCCATTGTGCTCGACAGCCTGGAACTTC 2587


RESULT  10
US-09-522-666-1/c
; Sequence 1, Application US/09522666
; Patent No. 6333167
; GENERAL INFORMATION:
;  APPLICANT: Shuey, David
;  APPLICANT: Quinet, Elaine
;  TITLE OF INVENTION: Methods and Reagents for Identifying Inhibitors of
;  TITLE OF INVENTION: Proteolysis of Membrane-Associated Proteins
;  FILE REFERENCE: 6-00
;  CURRENT APPLICATION NUMBER: US/09/522,666
;  CURRENT FILING DATE: 2000-03-10
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3092
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:contains
;   OTHER INFORMATION: APP-LacI fusion protein
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (12)..(3083)
US-09-522-666-1


  Query Match             47.6%;  Score 17.6;  DB 4;  Length 3092;
  Best Local Similarity   71.9%;  Pred. No. 74;
  Matches   23;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       4 ccggatcccatttttccttagcataacggaagt 35
           ||||||||  | ||| | ||||| ||  | ||
Db    1972 CCGGATCCTTTCTTCTTCAGCATCACCAAGGT 1941


RESULT  11
US-08-321-478-2
; Sequence 2, Application US/08321478
; Patent No. 5527677
;  GENERAL INFORMATION:
;     APPLICANT:  DEGUCHI, Takeo
;     APPLICANT:  KINOSHITA, Moritoshi
;     APPLICANT:  KATSURAGI, Kiyonori
;     APPLICANT:  SHIN, Sadahito
;     TITLE OF INVENTION:  HUMAN ARYLAMINE N-ACETYLTRANSFERASE
;     TITLE OF INVENTION:  GENES
;     NUMBER OF SEQUENCES:  13
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;       STREET:  2100 Pennsylvania Avenue, N.W.
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  United States
```

```
;       ZIP:  20037-3202
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/321,478
;       FILING DATE:  11-OCT-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/038,667
;       FILING DATE:  23-MAR-1993
;       APPLICATION NUMBER:  JP 64669/1992
;       FILING DATE:  23-MAR-1992
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 293-7060
;       TELEFAX:  (202) 293-7860
;       TELEX:  6491103
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6464 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  723..1595
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  717..1936
;     FEATURE:
;       NAME/KEY:  polyA_signal
;       LOCATION:  1794..1799
;     FEATURE:
;       NAME/KEY:  polyA_signal
;       LOCATION:  1800..1805
US-08-321-478-2


  Query Match           47.0%;  Score 17.4;  DB 1;  Length 6464;
  Best Local Similarity   77.8%;  Pred. No. 1.1e+02;
  Matches   21;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       3 gccggatcccattttccttagcataac 29
           || |||  ||||| |||||||||| | |
Db    5386 GCTGGAGGCCATTATCCTTAGCAAACC 5412


RESULT  12
US-08-321-478-4
; Sequence 4, Application US/08321478
; Patent No. 5527677
;  GENERAL INFORMATION:
;    APPLICANT:  DEGUCHI, Takeo
;    APPLICANT:  KINOSHITA, Moritoshi
;    APPLICANT:  KATSURAGI, Kiyonori
;    APPLICANT:  SHIN, Sadahito
;    TITLE OF INVENTION:  HUMAN ARYLAMINE N-ACETYLTRANSFERASE
;    TITLE OF INVENTION:  GENES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,478
;      FILING DATE:  11-OCT-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/038,667
;      FILING DATE:  23-MAR-1993
;      APPLICATION NUMBER:  JP 64669/1992
;      FILING DATE:  23-MAR-1992
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6464 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  723..1595
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  717..1936
;    FEATURE:
;      NAME/KEY:  polyA_signal
;      LOCATION:  1794..1799
;    FEATURE:
;      NAME/KEY:  polyA_signal
;      LOCATION:  1800..1805
US-08-321-478-4


  Query Match           47.0%;  Score 17.4;  DB 1;  Length 6464;
  Best Local Similarity   77.8%;  Pred. No. 1.1e+02;
  Matches   21;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       3 gccggatcccattttccttagcataac 29
           || |||  ||||| |||||||||| | |
Db    5386 GCTGGAGGCCATTATCCTTAGCAAACC 5412


RESULT  13
US-08-321-478-6
; Sequence 6, Application US/08321478
; Patent No. 5527677
;  GENERAL INFORMATION:
;    APPLICANT:  DEGUCHI, Takeo
;    APPLICANT:  KINOSHITA, Moritoshi
;    APPLICANT:  KATSURAGI, Kiyonori
;    APPLICANT:  SHIN, Sadahito
;    TITLE OF INVENTION:  HUMAN ARYLAMINE N-ACETYLTRANSFERASE
;    TITLE OF INVENTION:  GENES
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sughrue, Mion, Zinn, Macpeak & Seas
;      STREET:  2100 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  United States
;      ZIP:  20037-3202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,478
;      FILING DATE:  11-OCT-1994
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/038,667
;      FILING DATE:  23-MAR-1993
;      APPLICATION NUMBER:  JP 64669/1992
;      FILING DATE:  23-MAR-1992
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 293-7060
;      TELEFAX:  (202) 293-7860
;      TELEX:  6491103
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6464 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  723..1595
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  717..1936
;    FEATURE:
;      NAME/KEY:  polyA_signal
;      LOCATION:  1794..1799
;    FEATURE:
;      NAME/KEY:  polyA_signal
;      LOCATION:  1800..1805
US-08-321-478-6


  Query Match             47.0%;  Score 17.4;  DB 1;  Length 6464;
  Best Local Similarity   77.8%;  Pred. No. 1.1e+02;
  Matches   21;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       3 gccggatcccatttccttagcataac 29
           || |||  ||||| |||||||||| | |
Db    5386 GCTGGAGGCCATTATCCTTAGCAAACC 5412


RESULT  14
US-08-966-958-1/c
; Sequence 1, Application US/08966958
; Patent No. 5928908
;  GENERAL INFORMATION:
;    APPLICANT:  Dunn, John
;    APPLICANT:  Randesi, Matthew
;    TITLE OF INVENTION:  METHODS FOR INTRODUCING UNIDIRECTIONAL
;    TITLE OF INVENTION:  DELETIONS
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brookhaven National Laboratory
;      STREET:  P.O. Box 5000
;      CITY:  Upton
;      STATE:  New York
;      COUNTRY:  US
;      ZIP:  11973
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/966,958
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bogosian, Margaret
;      REGISTRATION NUMBER:  25,324
;      REFERENCE/DOCKET NUMBER:  AUI97-14
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (516) 344-3341
;      TELEFAX:  (516) 344-3729
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10754 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-966-958-1


  Query Match             47.0%;  Score 17.4;  DB 2;  Length 10754;
  Best Local Similarity   77.8%;  Pred. No. 1.2e+02;
  Matches   21;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcgccggatcccattttccttagcata 27
           | || |||  ||||| |||||||||| |
Db    3992 GAGCTGGAGGCCATTATCCTTAGCAAA 3966


RESULT  15
US-09-215-817-1/c
; Sequence 1, Application US/09215817
; Patent No. 5968786
;  GENERAL INFORMATION:
;    APPLICANT:  Dunn, John
;    APPLICANT:  Randesi, Matthew
;    TITLE OF INVENTION:  METHODS FOR INTRODUCING UNIDIRECTIONAL
;    TITLE OF INVENTION:  DELETIONS
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brookhaven National Laboratory
;      STREET:  P.O. Box 5000
;      CITY:  Upton
;      STATE:  New York
;      COUNTRY:  US
;      ZIP:  11973
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/215,817
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/966,958
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bogosian, Margaret
;      REGISTRATION NUMBER:  25,324
;      REFERENCE/DOCKET NUMBER:  AUI97-14
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (516) 344-3341
;      TELEFAX:  (516) 344-3729
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10754 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-215-817-1


  Query Match             47.0%;  Score 17.4;  DB 2;  Length 10754;
  Best Local Similarity   77.8%;  Pred. No. 1.2e+02;
  Matches   21;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 gcgccggatcccattttccttagcata 27
           | || |||  ||||| |||||||||| |
Db    3992 GAGCTGGAGGCCATTATCCTTAGCAAA 3966
```

Search completed: July 31, 2002, 19:33:02
Job time: 8478 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 20:58:36 ; Search time 337.68 Seconds
(without alignments)
188.124 Million cell updates/sec

Title: US-09-824-567-4
Perfect score: 37
Sequence: 1 gcgccggatcccatttttccttagcataacggaagtcc 37

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		37	100.0	37	22	AAD20240	Chlamydia pneumoni
c 2		25.6	69.2	1799	22	AAD20238	Chlamydia pneumoni
c 3		25.6	69.2	1230025	20	AAX91990	Nucleotide sequenc
c 4		20.8	56.2	273254	21	AAC81914	Chlamydia pneumoni
5		20.8	56.2	1230025	20	AAX91990	Nucleotide sequenc
c 6		20.6	55.7	238	21	AAC97319	Helicobacter pylor
c 7		20.6	55.7	448	21	AAC97248	Helicobacter pylor
c 8		20.6	55.7	474	21	AAC97239	Helicobacter pylor
c 9		20.6	55.7	3636	23	AAS53717	Helicobacter pylor
c 10		20	54.1	20	20	AAX92471	PCR primer used to
c 11		20	54.1	362	19	AAV38294	Human C-C chemokin
c 12		20	54.1	362	21	AAA47548	Primate CTACK nucl
c 13		20	54.1	439	21	AAZ56449	Human CC type chem
14		19.6	53.0	33	21	AAZ56459	Human interleukin
15		19.6	53.0	1246	24	ABA05830	A thaliana triacyl
c 16		19.6	53.0	10157	22	AAS46233	DNA encoding novel
17		19.6	53.0	269223	22	AAF28554	Genomic fragment #
18		19.4	52.4	395	23	ABL22791	Drosophila melanog
c 19		19.4	52.4	2270	23	ABL19206	Drosophila melanog
c 20		19.4	52.4	2499	23	ABL22790	Drosophila melanog
21		19.4	52.4	1664976	19	AAV21209	Methanococcus jann
22		19.2	51.9	253	20	AAX51467	Human secreted pro
c 23		19.2	51.9	13169	23	ABL18727	Drosophila melanog
24		19.2	51.9	18177	10	AAN90490	DNA of human retin
25		19.2	51.9	18303	20	AAX04502	Human retinoblasto
c 26		19	51.4	388	22	AAI82264	Human polynucleoti
27		19	51.4	27681	22	AAS36497	Human cardiovascul
28		19	51.4	27681	22	AAS36498	Human cardiovascul
29		19	51.4	27681	22	AAK85843	Human immune/haema
30		18.8	50.8	121	22	ABA77671	Retinoblastoma mut
c 31		18.8	50.8	121	22	ABA77672	Retinoblastoma mut
c 32		18.8	50.8	431	22	AAI82511	Human polynucleoti
33		18.8	50.8	9652	23	ABL18818	Drosophila melanog
c 34		18.8	50.8	35832	23	ABL18726	Drosophila melanog
c 35		18.6	50.3	343	22	AAL25392	Human breast cance
c 36		18.6	50.3	353	22	AAL07646	Human breast cance
c 37		18.6	50.3	354	22	AAL16549	Human breast cance
c 38		18.6	50.3	364	21	AAC01669	Human secreted pro
c 39		18.6	50.3	795	22	AAL26493	Human breast cance
c 40		18.6	50.3	1424	21	AAC49703	Arabidopsis thalia
c 41		18.6	50.3	1428	21	AAC40085	Arabidopsis thalia
c 42		18.6	50.3	1586	22	AAD08981	Zea mays mutM homo
43		18.6	50.3	1975	19	AAV43000	Streptococcus pneu
c 44		18.6	50.3	2214	21	AAA48493	Streptococcus pneu
c 45		18.6	50.3	2214	23	AAS55964	Streptococcus pneu

ALIGNMENTS

RESULT 1
AAD20240
ID AAD20240 standard; DNA; 37 BP.
XX
AC AAD20240;
XX
DT 15-JAN-2002 (first entry)
XX
DE Chlamydia pneumoniae ATP-binding cassette gene amplifying 3'PCR primer.
XX
KW ATP-binding cassette; antibiotic; vaccine; infection; therapy; poxvirus;
KW PCR primer; ss.
XX
OS Chlamydia pneumoniae.
XX
PN WO200174863-A2.
XX
PD 11-OCT-2001.
XX
PF 04-APR-2001; 2001WO-CA00455.
XX
PR 04-APR-2000; 2000US-194464P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI; 2001-648549/74.
XX
PT Novel Chlamydia ATP-binding cassette and corresponding DNA molecule for
PT preventing, diagnosing and treating Chlamydia infections in mammals, in
PT particular humans -

```
XX
PS   Claim 41; Page 53; 88pp; English.
XX
CC   The present invention relates to novel Chlamydia pneumoniae ATP-binding
CC   cassette protein and its corresponding gene. Sequences of the invention
CC   are useful for detecting Chlamydia infection by assaying a body fluid
CC   of a mammal with the components. They are also used as vaccines. ATP
CC   binding cassette antibodies and vaccines of the invention are useful
CC   for preventing or treating Chlamydia infection e.g. infection caused
CC   by C. trachomatis, C. psittaci, C. pneumoniae or C. pecorum in mammals,
CC   such as humans. The nucleic acid molecules are useful for producing
CC   ATP-binding cassettes, in the construction of vaccine vectors such
CC   as poxviruses, which are further useful for preventing and/or treating
CC   Chlamydia infection and in the construction of attenuated Chlamydia
CC   strains that can over-express the nucleic acid molecules or express
CC   it in a non-toxic, mutated form. The present DNA sequence is a 3' PCR
CC   primer which is used for amplifying Chlamydia pneumoniae ATP-binding
CC   cassette DNA.
XX
SQ   Sequence 37 BP; 8 A; 12 C; 8 G; 9 T; 0 other;

  Query Match           100.0%;  Score 37;  DB 22;  Length 37;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-07;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gcgccggatcccattttccttagcataacggaagtcc 37
           |||||||||||||||||||||||||||||||||||||
Db       1 gcgccggatcccattttccttagcataacggaagtcc 37


RESULT   2
AAD20238/c
ID   AAD20238 standard; DNA; 1799 BP.
XX
AC   AAD20238;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Chlamydia pneumoniae ATP-binding cassette gene.
XX
KW   ATP-binding cassette; antibiotic; vaccine; infection; therapy; poxvirus;
KW   ds.
XX
OS   Chlamydia pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   CDS             101..1699
FT                   /*tag=  a
FT                   /product= "ATP-binding cassette protein"
XX
PN   WO200174863-A2.
XX
PD   11-OCT-2001.
XX
PF   04-APR-2001; 2001WO-CA00455.
XX
PR   04-APR-2000; 2000US-194464P.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
XX
PI   Murdin AD,  Oomen RP,  Wang J,  Dunn P;
XX
DR   WPI; 2001-648549/74.
DR   P-PSDB; AAE12212.
XX
PT   Novel Chlamydia ATP-binding cassette and corresponding DNA molecule for
PT   preventing, diagnosing and treating Chlamydia infections in mammals, in
PT   particular humans -
XX
PS   Claim 2; Fig 1; 88pp; English.
XX
CC   The present invention relates to novel Chlamydia pneumoniae ATP-binding
CC   cassette protein and its corresponding gene. Sequences of the invention
CC   are useful for detecting Chlamydia infection by assaying a body fluid
CC   of a mammal with the components. They are also used as vaccines. ATP
CC   binding cassette antibodies and vaccines of the invention are useful
CC   for preventing or treating Chlamydia infection e.g. infection caused
CC   by C. trachomatis, C. psittaci, C. pneumoniae or C. pecorum in mammals,
CC   such as humans. The nucleic acid molecules are useful for producing
CC   ATP-binding cassettes, in the construction of vaccine vectors such
CC   as poxviruses, which are further useful for preventing and/or treating
CC   Chlamydia infection and in the construction of attenuated Chlamydia
CC   strains that can over-express the nucleic acid molecules or express
CC   it in a non-toxic, mutated form. The present sequence is a gene encoding
CC   Chlamydia pneumoniae ATP-binding cassette.
XX
SQ   Sequence 1799 BP; 560 A; 439 C; 294 G; 506 T; 0 other;

  Query Match           69.2%;  Score 25.6;  DB 22;  Length 1799;
  Best Local Similarity  87.5%;  Pred. No. 0.11;
  Matches   28;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       6 ggatcccatttttccttagcataacggaagtcc 37
           ||  |  ||||||||||||||||||||||||||||
Db    1703 GGTGCTAATTTTCCTTAGCATAACGGAAGTCC 1672


RESULT   3
AAX91990/c
ID   AAX91990 standard; DNA; 1230025 BP.
XX
AC   AAX91990;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope; ss.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Claim 1; Page 291-611; 1912pp; English.
XX
CC   The present sequence represents the complete genome of Chlamydia
CC   pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
CC   respiratory disease such as pneumonia and bronchitis and is thought
CC   to be a contributing factor in heart disease, sarcoidosis, sinusitis,
CC   purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
CC   encoded by the open reading frames of the C. pneumoniae genome (see
CC   AAY34584-Y35879) can be used in immunogenic compositions as vaccines.
CC   Vectors containing C. pneumoniae nucleotides sequences can also be
CC   used as immunogenic compositions, especially where the vector directs
CC   the expression of a neutralising epitope of C. pneumoniae.
XX
SQ   Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;
```

```
  Query Match           69.2%;  Score 25.6;  DB 20;  Length 1230025;
  Best Local Similarity  87.5%;  Pred. No. 0.45;
  Matches   28;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      6 ggatcccatttttccttagcataacggaagtcc 37
          ||  |  ||||||||||||||||||||||||||||
Db 246083 GGTGCTAATTTTCCTTAGCATAACGGAAGTCC 246052


RESULT   4
AAC81914/c
ID   AAC81914 standard; DNA; 273254 BP.
XX
AC   AAC81914;
XX
DT   27-FEB-2001  (first entry)
XX
DE   Chlamydia pneumoniae genome DNA.
XX
KW   Genome; diagnosis; vaccine; ds.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO200027994-A2.
XX
PD   18-MAY-2000.
XX
PF   12-NOV-1999;   99WO-US26923.
XX
PR   12-NOV-1998;   98US-0108279.
PR   08-APR-1999;   99US-0128606.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Stephens R,  Mitchell W,  Kalman S,  Davis R;
XX
DR   WPI; 2000-376516/32.
XX
PT   Isolated nucleic acid for use in diagnostic and analytical methods
PT   encodes genomic sequence of Chlamydia pneumoniae -
XX
PS   Claim 2; Page 128-320; 320pp; English.
XX
CC   This invention describes a novel nucleic acid (N1) encoding a Chlamydia
CC   pneumoniae protein (P1), given in the specification. The isolated nucleic
CC   acid is useful for diagnostic and analytical methods, such as,
CC   hybridization-based assays or amplification-based assays. The protein may
CC   be used for diagnostic purposes, for their enzymatic or structural
CC   activity, or as a vaccine. The invention also describes (1) a probe
CC   comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
CC   (N2) that hybridizes under stringent conditions to N1; (3) an expression
CC   cassette comprising N1 under the transcriptional regulation of a
CC   transcriptional initiation region functional in an expression host, and a
CC   transcriptional termination region; (4) a cell comprising an expression
CC   cassette of (3) as part of an extrachromosomal element or integrated into
CC   the genome of a host cell as a result of induction of the expression
CC   cassette into the host cell, and the cellular progeny of the host cell;
CC   (5) a method for producing a P1 comprising growing a cell of (4) where
CC   the protein is expressed and isolating the protein free of other
CC   proteins; (6) a purified polypeptide composition comprising at least 50
CC   weight % of P1; and (7) a monoclonal antibody binding specifically to the
CC   peptide of (6).
XX
SQ   Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;


  Query Match           56.2%;  Score 20.8;  DB 21;  Length 273254;
  Best Local Similarity  78.1%;  Pred. No. 47;
  Matches   25;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      6 ggatcccatttttccttagcataacggaagtcc 37
            |||||  |  ||||||||||||  | |  |||| |
Db 152997 GGATCTCTCTTTCCTTAGCATTAGGAAAGTAC 152966


RESULT   5
AAX91990
ID   AAX91990 standard; DNA; 1230025 BP.
XX
AC   AAX91990;
XX
DT   13-SEP-1999  (first entry)
XX
DE   Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope; ss.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO9927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;   98WO-IB01890.
XX
PR   04-NOV-1998;   98US-0107078.
PR   21-NOV-1997;   97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30.
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Claim 1; Page 291-611; 1912pp; English.
XX
CC   The present sequence represents the complete genome of Chlamydia
CC   pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
CC   respiratory disease such as pneumonia and bronchitis and is thought
CC   to be a contributing factor in heart disease, sarcoidosis, sinusitis,
CC   purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
CC   encoded by the open reading frames of the C. pneumoniae genome (see
CC   AAY34584-Y35879) can be used in immunogenic compositions as vaccines.
CC   Vectors containing C. pneumoniae nucleotides sequences can also be
CC   used as immunogenic compositions, especially where the vector directs
CC   the expression of a neutralising epitope of C. pneumoniae.
XX
SQ   Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;


  Query Match           56.2%;  Score 20.8;  DB 20;  Length 1230025;
  Best Local Similarity  78.1%;  Pred. No. 63;
  Matches   25;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      6 ggatcccatttttccttagcataacggaagtcc 37
            |||||  |  ||||||||||||  | |  |||| |
Db 463794 ggatctctctttccttagcattaggaaagtac 463825


RESULT   6
AAC97319/c
ID   AAC97319 standard; DNA; 238 BP.
XX
AC   AAC97319;
XX
DT   23-FEB-2001  (first entry)
XX
DE   Helicobacter pylori bait polypeptide nucleotide sequence #91.
XX
KW   Helicobacter pylori; two-hybrid system; protein-protein interaction;
```

KW antibacterial; bait polypeptide; gastric ulcer; ds.
XX
OS Helicobacter pylori.
XX
PN WO200066722-A1.
XX
PD 09-NOV-2000.
XX
PF 14-APR-2000; 2000WO-IB00603.
XX
PR 30-APR-1999; 99EP-0401066.
XX
PA (HYBR-) HYBRIGENICS SA.
XX
PI Legrain P, Selig L, Rain J;
XX
DR WPI; 2000-687535/67.
DR P-PSDB; AAB52573.
XX
PT A two-hybrid system for identifying compounds useful in the treatment
PT of e.g. gastric ulcers comprises producing a collection of recombinant
PT cell clones -
XX
PS Example 5; Page 197; 267pp; English.
XX
CC The present sequence encodes a bait polypeptide used in a Helicobacter
CC pylori two-hybrid screen to identify protein-protein interactions.
CC The method is used to identify a recombinant cell clone expressing a
CC prey polypeptide which is capable of interacting with the bait
CC polypeptide. The two hybrid system is useful for screening compounds
CC for antibacterial activity. It may be used in the treatment of gastric
CC ulcers. The polynucleotides are useful as amplification primers or
CC specific detection probes. The polypeptides, vectors or host cells can
CC be used as immunogens to produce mono- or polyclonal antibodies. The
CC polynucleotides, polypeptides, antibodies, vectors, host cells or
CC modulating agents can be used to produce a pharmaceutical composition.
XX
SQ Sequence 238 BP; 66 A; 40 C; 64 G; 68 T; 0 other;

Query Match 55.7%; Score 20.6; DB 21; Length 238;
Best Local Similarity 85.2%; Pred. No. 13;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 ggatcccattttccttagcataacgga 32
| || || ||||||||||||||||||||
Db 103 GAATGCCCATTTCCTTAGCATAACGGA 77

RESULT 7
AAC97248/c
ID AAC97248 standard; DNA; 448 BP.
XX
AC AAC97248;
XX
DT 23-FEB-2001 (first entry)
XX
DE Helicobacter pylori bait polypeptide nucleotide sequence #20.
XX
KW Helicobacter pylori; two-hybrid system; protein-protein interaction;
KW antibacterial; bait polypeptide; gastric ulcer; ds.
XX
OS Helicobacter pylori.
XX
PN WO200066722-A1.
XX
PD 09-NOV-2000.
XX
PF 14-APR-2000; 2000WO-IB00603.
XX
PR 30-APR-1999; 99EP-0401066.
XX
PA (HYBR-) HYBRIGENICS SA.
XX
PI Legrain P, Selig L, Rain J;
XX
DR WPI; 2000-687535/67.
DR P-PSDB; AAB52502.
XX
PT A two-hybrid system for identifying compounds useful in the treatment
PT of e.g. gastric ulcers comprises producing a collection of recombinant
PT cell clones -
XX
PS Example 5; Page 102-103; 267pp; English.
XX
CC The present sequence encodes a bait polypeptide used in a Helicobacter
CC pylori two-hybrid screen to identify protein-protein interactions.
CC The method is used to identify a recombinant cell clone expressing a
CC prey polypeptide which is capable of interacting with the bait
CC polypeptide. The two hybrid system is useful for screening compounds
CC for antibacterial activity. It may be used in the treatment of gastric
CC ulcers. The polynucleotides are useful as amplification primers or
CC specific detection probes. The polypeptides, vectors or host cells can
CC be used as immunogens to produce mono- or polyclonal antibodies. The
CC polynucleotides, polypeptides, antibodies, vectors, host cells or
CC modulating agents can be used to produce a pharmaceutical composition.
XX
SQ Sequence 448 BP; 143 A; 81 C; 102 G; 122 T; 0 other;

Query Match 55.7%; Score 20.6; DB 21; Length 448;
Best Local Similarity 85.2%; Pred. No. 15;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 ggatcccattttccttagcataacgga 32
| || || ||||||||||||||||||||
Db 352 GAATGCCCATTTCCTTAGCATAACGGA 326

RESULT 8
AAC97239/c
ID AAC97239 standard; DNA; 474 BP.
XX
AC AAC97239;
XX
DT 23-FEB-2001 (first entry)
XX
DE Helicobacter pylori bait polypeptide nucleotide sequence #11.
XX
KW Helicobacter pylori; two-hybrid system; protein-protein interaction;
KW antibacterial; bait polypeptide; gastric ulcer; ds.
XX
OS Helicobacter pylori.
XX
PN WO200066722-A1.
XX
PD 09-NOV-2000.
XX
PF 14-APR-2000; 2000WO-IB00603.
XX
PR 30-APR-1999; 99EP-0401066.
XX
PA (HYBR-) HYBRIGENICS SA.
XX
PI Legrain P, Selig L, Rain J;
XX
DR WPI; 2000-687535/67.
DR P-PSDB; AAB52493.
XX
PT A two-hybrid system for identifying compounds useful in the treatment
PT of e.g. gastric ulcers comprises producing a collection of recombinant
PT cell clones -
XX
PS Example 5; Page 92-93; 267pp; English.
XX
CC The present sequence encodes a bait polypeptide used in a Helicobacter

CC pylori two-hybrid screen to identify protein-protein interactions.
CC The method is used to identify a recombinant cell clone expressing a
CC prey polypeptide which is capable of interacting with the bait
CC polypeptide. The two hybrid system is useful for screening compounds
CC for antibacterial activity. It may be used in the treatment of gastric
CC ulcers. The polynucleotides are useful as amplification primers or
CC specific detection probes. The polypeptides, vectors or host cells can
CC be used as immunogens to produce mono- or polyclonal antibodies. The
CC polynucleotides, polypeptides, antibodies, vectors, host cells or
CC modulating agents can be used to produce a pharmaceutical composition.
XX
SQ Sequence 474 BP; 164 A; 91 C; 104 G; 115 T; 0 other;

Query Match 55.7%; Score 20.6; DB 21; Length 474;
Best Local Similarity 85.2%; Pred. No. 15;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 ggatcccatttccttagcataacgga 32
| || || ||||||||||||||||||||
Db 472 GAATGCCCATTTCCTTAGCATAACGGA 446

RESULT 9
AAS53717/c
ID AAS53717 standard; DNA; 3636 BP.
XX
AC AAS53717;
XX
DT 13-FEB-2002 (first entry)
XX
DE Helicobacter pylori DNA for cellular proliferation protein #171.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Helicobacter pylori.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU35858.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 7354; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3636 BP; 1184 A; 633 C; 873 G; 946 T; 0 other;

Query Match 55.7%; Score 20.6; DB 23; Length 3636;
Best Local Similarity 85.2%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 ggatcccatttccttagcataacgga 32
| || || ||||||||||||||||||||
Db 1141 GAATGCCCATTTCCTTAGCATAACGGA 1115

RESULT 10
AAX92471/c
ID AAX92471 standard; DNA; 20 BP.
XX
AC AAX92471;
XX
DT 13-SEP-1999 (first entry)
XX
DE PCR primer used to amplify an ORF of Chlamydia pneumoniae.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; PCR primer; ss.
XX
OS Synthetic.
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1514; Disclosure; 1912pp; English.
XX
CC AAX91991-X97517 represent PCR primers used to amplify open reading
CC frames and other nucleic acid sequences from the genome of
CC Chlamydia pneumoniae (see AAX91990). C. pneumoniae causes respiratory
CC disease such as pneumonia and bronchitis and is thought to be a
CC contributing factor in heart disease, sarcoidosis, sinusitis, purulent
CC otitis media, erythema nodosum or pharyngitis. The polypeptides encoded
CC by the open reading frames of the C. pneumoniae genome (see AAY34584-
CC AAY35879) can be used in immunogenic compositions as vaccines. Vectors
CC containing C. pneumoniae nucleotides sequences can also be used as
CC immunogenic compositions, especially where the vector directs the
CC expression of a neutralising epitope of C. pneumoniae.
XX
SQ Sequence 20 BP; 4 A; 4 C; 6 G; 6 T; 0 other;

```
  Query Match             54.1%;  Score 20;  DB 20;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      18 ccttagcataacggaagtcc 37
           ||||||||||||||||||||
Db      20 CCTTAGCATAACGGAAGTCC 1


RESULT  11
AAV38294/c
ID   AAV38294 standard; cDNA; 362 BP.
XX
AC   AAV38294;
XX
DT   12-OCT-1998  (first entry)
XX
DE   Human C-C chemokine DGWCC cDNA.
XX
KW   DGWCC; DNAX groin wound expressed CC chemokine; cytokine; human;
KW   immune system; cancer; cell proliferation; therapy; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..339
FT                   /*tag=  a
FT   sig_peptide     1..72
FT                   /*tag=  b
FT   mat_peptide     73..336
FT                   /*tag=  c
XX
PN   WO9823750-A2.
XX
PD   04-JUN-1998.
XX
PF   26-NOV-1997;   97WO-US21092.
XX
PR   05-DEC-1996;   96US-0761071.
PR   27-NOV-1996;   96US-0031805.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Hedrick JA,  Morales J,  Vicari A,  Zlotnik A;
XX
DR   WPI; 1998-322730/28.
DR   P-PSDB; AAW60650.
XX
PT   DVic-1 and DGWCC chemokines  - useful for developing products for
PT   treating abnormal physiology or development, e.g. cancerous or
PT   degenerative conditions
XX
PS   Disclosure; Page 62; 71pp; English.
XX
CC   This cDNA sequence codes for novel human DNAX groin wound expressed
CC   CC chemokine (DGWCC) (see AAW60650).  DGWCC cDNA can be obtained from
CC   e.g. skin, epithelial or wound healing libraries by PCR
CC   amplification or by hybridisation.  Also disclosed is novel human
CC   DNAX Vic-1 (DVia-1) (see AAW60649), as well as expression vectors and
CC   host cells.  DGWCC and DVic-1 play a role in the regulation or
CC   development of neuronal or haematopoietic cells, e.g. lymphoid
CC   cells, which affect immunological responses.  They can be used in
CC   the treatment of conditions associated with abnormal physiology or
CC   development, including abnormal proliferation, e.g. cancerous
CC   conditions or degenerative conditions.  Abnormal proliferation,
CC   regeneration, degeneration, and atrophy may be modulated by
CC   appropriate therapeutic treatment using products of the invention.
CC   The products can also be used for detection, diagnosis and drug
CC   screening.
XX
SQ   Sequence 362 BP; 91 A; 116 C; 85 G; 70 T; 0 other;

  Query Match             54.1%;  Score 20;  DB 19;  Length 362;
  Best Local Similarity   72.2%;  Pred. No. 27;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 gcgccggatcccatttttccttagcataacggaagtc 36
           | ||   | ||||||||||||||||  |  || ||
Db     344 GGGCTTCAGCCCATTTTCCTTAGCATCCCAAAATTC 309


RESULT  12
AAA47548/c
ID   AAA47548 standard; DNA; 362 BP.
XX
AC   AAA47548;
XX
DT   20-OCT-2000  (first entry)
XX
DE   Primate CTACK nucleotide sequence.
XX
KW   Cutaneous T-cell attracting chemokine; CTACK; skin; cell movement;
KW   migration; vasoactive intestinal contractor; Vic; GPR2; agonist;
KW   antagonist; antibody; immunological condition; mutein; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..339
FT                   /*tag=  a
FT                   /product= CTACK
FT   sig_peptide     1..72
FT                   /*tag=  b
FT   mat_peptide     73..336
FT                   /*tag=  c
XX
PN   WO200038713-A1.
XX
PD   06-JUL-2000.
XX
PF   23-DEC-1999;   99WO-US30819.
XX
PR   24-DEC-1998;   98US-0113858.
PR   27-MAY-1999;   99US-0322580.
XX
PA   (SCHE ) SCHERING CORP.
XX
PI   Wang W,  Oldham ER,  Soto H,  Lui Y,  Hudak SA,  Homey B,  Morales JM;
PI   Kellermann S,  McEvoy LM,  Zlotnik A;
XX
DR   WPI; 2000-465633/40.
DR   P-PSDB; AAB01453.
XX
PT   Modulating cell movement within the skin, useful for treating
PT   immunological skin conditions or diseases comprises administering T
PT   cell-attracting chemokine or vasoactive intestinal contractor chemokine
PT   agonists or antagonists
XX
PS   Example 3; Page 73; 79pp; English.
XX
CC   Modulating movement of a cell within or to the skin of a mammal can
CC   be achieved by administering an antagonist or agonist of cutaneous T
CC   cell-attracting chemokine (CTACK) or vasoactive intestinal contractor
CC   (Vic) chemokine. The antagonist is selected from a mutein of natural
CC   CTACK or Vic, an antibody which neutralises CTACK or Vic or an
CC   antibody which block GPR2 ligand binding. The CTACK or Vic agonists
CC   or antagonists are useful for treating medical conditions or diseases
CC   associated with immunological conditions of the skin.
XX
SQ   Sequence 362 BP; 91 A; 116 C; 85 G; 70 T; 0 other;

  Query Match             54.1%;  Score 20;  DB 21;  Length 362;
```

```
  Best Local Similarity   72.2%;  Pred. No. 27;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 gcgccggatcccattttccttagcataacggaagtc 36
          | ||   | ||||||||||||||||||  |  || ||
Db    344 GGGCTTCAGCCCATTTTCCTTAGCATCCCAAAATTC 309


RESULT  13
AAZ56449/c
ID   AAZ56449 standard; cDNA to mRNA; 439 BP.
XX
AC   AAZ56449;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Human CC type chemokine interleukin C encoding cDNA.
XX
KW   Human; CC type chemokine; IL-C; cancer; infectious disease;
KW   interleukin C; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             60..398
FT                   /*tag=  a
FT                   /product= "Interleukin C protein"
FT                   /note= "a CC type chemokine with homology to CC type
FT                          chemokine MC148R in poxvirus MCV (Molluscum
FT                          contagiosum virus)"
XX
PN   JP11302298-A.
XX
PD   02-NOV-1999.
XX
PF   20-APR-1998;   98JP-0109434.
XX
PR   20-APR-1998;   98JP-0109434.
XX
PA   (SHIO ) SHIONOGI & CO LTD.
XX
DR   WPI; 2000-075344/07.
DR   P-PSDB; AAY57771.
XX
PT   Human CC type chemokine Interleukin (IL)-C - for the treatment of
PT   cancers and infectious diseases
XX
PS   Claim 3; Page 11-12; 15pp; Japanese.
XX
CC   The present sequence encodes a human CC type chemokine interleukin C
CC   (IL-C) protein which has amino acid homology to CC type chemokine
CC   MC148R in poxvirus MCV ( Molluscum contagiosum virus ) and is expressed
CC   structurally mainly in a thymus and a placenta. The polynucleotide
CC   encoding human IL-C is useful for the treatment of cancers, infectious
CC   diseases and diseases accompanied by abnormality in the structure and
CC   expression of the IL-C gene.
XX
SQ   Sequence 439 BP; 121 A; 126 C; 113 G; 79 T; 0 other;


  Query Match             54.1%;  Score 20;  DB 21;  Length 439;
  Best Local Similarity   72.2%;  Pred. No. 28;
  Matches   26;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 gcgccggatcccattttccttagcataacggaagtc 36
          | ||   | ||||||||||||||||||  |  || ||
Db    403 GGGCTTCAGCCCATTTTCCTTAGCATCCCAAAATTC 368


RESULT  14
AAZ56459
ID   AAZ56459 standard; DNA; 33 BP.
XX
AC   AAZ56459;
XX
DT   17-MAR-2000  (first entry)
XX
DE   Human interleukin C PCR primer hILC-C1 SEQ ID NO:11.
XX
KW   Human; CC type chemokine; IL-C; cancer; infectious disease;
KW   interleukin C; PCR primer; ss.
XX
OS   Homo sapiens.
XX
PN   JP11302298-A.
XX
PD   02-NOV-1999.
XX
PF   20-APR-1998;   98JP-0109434.
XX
PR   20-APR-1998;   98JP-0109434.
XX
PA   (SHIO ) SHIONOGI & CO LTD.
XX
DR   WPI; 2000-075344/07.
XX
PT   Human CC type chemokine Interleukin (IL)-C - for the treatment of
PT   cancers and infectious diseases
XX
PS   Example 3; Page 9; 15pp; Japanese.
XX
CC   The present invention describes a human CC type chemokine interleukin C
CC   (IL-C) protein which has amino acid homology to CC type chemokine
CC   MC148R in poxvirus MCV ( Molluscum contagiosum virus ) and is expressed
CC   structurally mainly in a thymus and a placenta. The polynucleotide
CC   encoding human IL-C is useful for the treatment of cancers, infectious
CC   diseases and diseases accompanied by abnormality in the structure and
CC   expression of the IL-C gene. The present sequence represents a PCR
CC   primer for human IL-C, which is used in an example from the present
CC   invention.
XX
SQ   Sequence 33 BP; 8 A; 12 C; 4 G; 9 T; 0 other;


  Query Match             53.0%;  Score 19.6;  DB 21;  Length 33;
  Best Local Similarity   84.6%;  Pred. No. 24;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 gcgccggatcccattttccttagcat 26
          || |  || ||||||||||||||||||
Db      2 gctctagagcccattttccttagcat 27


RESULT  15
ABA05830
ID   ABA05830 standard; cDNA; 1246 BP.
XX
AC   ABA05830;
XX
DT   15-MAR-2002  (first entry)
XX
DE   A thaliana triacylglycerol lipase TAG-lipase 1 coding sequence.
XX
KW   Triacylglycerol lipase; TAG-lipase 1; acylhydrolase; transgenic plant;
KW   nutrition; polyene fatty acid; PEFA; plant oil; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   DE10026845-A1.
XX
PD   06-DEC-2001.
XX
PF   31-MAY-2000; 2000DE-1026845.
XX
PR   31-MAY-2000; 2000DE-1026845.
XX
```

PA (IPKP-) IPK INST PFLANZENGENETIK & KULTURPFLANZE.
XX
PI Koerner M, Berndt E, Fritsche K, Feussner I;
XX
DR WPI; 2002-098656/14.
XX
PT New polynucleotide, useful for producing transgenic food plants with
PT altered contents of polyene fatty acids, comprises recombinant nucleic
PT acid encoding plant acylhydrolase -
XX
PS Claim 5; Page 12; 25pp; German.
XX
CC The present invention provides coding sequences and proteins from
CC Arabidopsis thaliana which act as acylhydrolase enzymes. These are
CC designated triacylglycerol lipase (TAG-lipase) 1,2 and 3. The sequences
CC are used to produce transgenic plants, which are useful in human or
CC animal nutrition and have altered content of (oxygenated) polyene fatty
CC acids (PEFA) in the seed oil. They can also be used to identify and
CC isolate other TAG-lipase encoding sequences from genomic/cDNA libraries.
CC The present sequence is the TAG-lipase 1 coding sequence of the
CC invention.
CC Note: The present sequence is stated in the specification as encoding the
CC protein shown in SEQ ID NO: 2 (ABB04482). However, this is not the case.
XX
SQ Sequence 1246 BP; 324 A; 263 C; 296 G; 363 T; 0 other;

Query Match 53.0%; Score 19.6; DB 24; Length 1246;
Best Local Similarity 84.6%; Pred. No. 52;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       9 tcccatttccttagcataacggaag 34
           | | |||||||| ||||||| |||||
Db     838 ttcaattttcctaagcataatggaag 863
```

Search completed: July 31, 2002, 20:59:28
Job time: 7759 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 19:31:22 ; Search time 1882.78 Seconds
(without alignments)
411.244 Million cell updates/sec

Title: US-09-824-567-4
Perfect score: 37
Sequence: 1 gcgccggatcccattttccttagcataacggaagtcc 37

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	100.0	37	6	AX268344	AX268344 Sequence
c 2	25.6	69.2	1799	6	AX268341	AX268341 Sequence
c 3	25.6	69.2	11648	1	AE001606	AE001606 Chlamydia
4	25.6	69.2	11764	1	AE002216	AE002216 Chlamydop
c 5	25.6	69.2	299650	1	AP002545	AP002545 Chlamydop
c 6	25	67.6	1599	6	AX349501	AX349501 Sequence
c 7	21.6	58.4	166919	2	AL591675	AL591675 Mus muscu
8	21.6	58.4	187334	2	AL626766	AL626766 Mus muscu
c 9	21.6	58.4	226999	2	AC105488	AC105488 Rattus no
c 10	21.6	58.4	239486	2	AC097752	AC097752 Rattus no
11	21.4	57.8	151321	9	AC069483	AC069483 Homo sapi
12	21.4	57.8	170257	9	AC078909	AC078909 Homo sapi
13	21.2	57.3	70148	9	AL357061	AL357061 Human DNA
c 14	21.2	57.3	107967	9	AL353701	AL353701 Human DNA
c 15	21.2	57.3	339681	1	AP003009	AP003009 Mesorhizo
c 16	21	56.8	289	3	AF029813	AF029813 Hysterole
c 17	21	56.8	50829	2	AC017759	AC017759 Drosophil
c 18	21	56.8	140167	9	AC004010	AC004010 Homo sapi
19	21	56.8	153048	3	AC007891	AC007891 Drosophil
c 20	21	56.8	164317	9	AL390029	AL390029 Human DNA
21	21	56.8	171374	9	AC010980	AC010980 Homo sapi
c 22	21	56.8	179150	2	AC099634	AC099634 Mus muscu
23	21	56.8	179937	9	AC016683	AC016683 Homo sapi
24	21	56.8	204173	3	AC007814	AC007814 Drosophil
c 25	21	56.8	225432	9	AF027390	AF027390 Homo sapi
26	21	56.8	232380	3	AE003724	AE003724 Drosophil
27	20.8	56.2	637	9	HSA335036	AJ335036 Homo sapi
28	20.8	56.2	14917	1	AE001624	AE001624 Chlamydia
c 29	20.8	56.2	20415	1	AE002196	AE002196 Chlamydop
30	20.8	56.2	30734	8	AC079679	AC079679 Arabidops
31	20.8	56.2	65107	2	AC079962	AC079962 Homo sapi
c 32	20.8	56.2	126792	9	AC008102	AC008102 Homo sapi
33	20.8	56.2	138020	9	AC108043	AC108043 Homo sapi
34	20.8	56.2	165282	2	AP000822	AP000822 Homo sapi
c 35	20.8	56.2	170940	2	AC025132	AC025132 Homo sapi
36	20.8	56.2	174130	2	AC108026	AC108026 Homo sapi
37	20.8	56.2	195251	30	AC023862	AC023862 Homo sapi
38	20.8	56.2	210647	2	AC093021	AC093021 Mus muscu
c 39	20.8	56.2	232951	2	AC091322	AC091322 Mus muscu
40	20.8	56.2	300650	1	AP002546	AP002546 Chlamydop
c 41	20.6	55.7	238	6	AX044720	AX044720 Sequence
c 42	20.6	55.7	448	6	AX044578	AX044578 Sequence
c 43	20.6	55.7	474	6	AX044560	AX044560 Sequence
c 44	20.6	55.7	1237	8	AF117307	AF117307 Daucus ca
c 45	20.6	55.7	8005	8	DCA18706	Y18706 Daucus caro

ALIGNMENTS

RESULT 1
AX268344
LOCUS AX268344 37 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 4 from Patent WO0174863.
ACCESSION AX268344
VERSION AX268344.1 GI:16541565
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (sites)
AUTHORS Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
TITLE Chlamydia antigens and corresponding dna fragments and uses thereof
JOURNAL Patent: WO 0174863-A 4 11-OCT-2001;
Aventis Pasteur Limited (CA)
FEATURES Location/Qualifiers
source 1. .37
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="3' PCR primer"
BASE COUNT 8 a 12 c 8 g 9 t
ORIGIN

```
  Query Match           100.0%;  Score 37;  DB 6;  Length 37;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-05;
  Matches   37;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gcgccggatcccatttttccttagcataacggaagtcc 37
          |||||||||||||||||||||||||||||||||||||||
Db      1 GCGCCGGATCCCATTTTCCTTAGCATAACGGAAGTCC 37


RESULT   2
AX268341/c
LOCUS       AX268341               1799 bp    DNA     linear   PAT 29-OCT-2001
DEFINITION  Sequence 1 from Patent WO0174863.
ACCESSION   AX268341
VERSION     AX268341.1  GI:16541562
KEYWORDS    .
SOURCE      Chlamydophila pneumoniae.
  ORGANISM  Chlamydophila pneumoniae
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
REFERENCE   1  (sites)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0174863-A 1 11-OCT-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .1799
                     /organism="Chlamydophila pneumoniae"
                     /db_xref="taxon:83558"
     CDS             101. .1699
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD10302.1"
                     /db_xref="GI:16541563"
                     /translation="MRKISVGICITILLSLSVVLQGCKESSHSSTSRGELAINIRDEP
                     RSLDPRQVRLLSEISLVKHIYEGLVQENNLSGNIEPALAEDYSLSSDGLTYTFKLKSA
                     FWSNGDPLTAEDFIESWKQVATQEVSGIYAFALNPIKNVRKIQEGHLSIDHFGVHSPN
                     ESTLVVTLESPTSHFLKLLALPVFFPVHKSQRTLQSKSLPIASGAFYPKNIKQKQWIK
                     LSKNPHYYNQSQVETKTITIHFIPDANTAAKLFNQGKLNWQGPPWGERIPQETLSNLQ
                     SKGHLHSFDVAGTSWLTFNINKFPLNNMKLREALASALDKEALVSTIFLGRAKTADHL
                     LPTNIHSYPEHQKQEMAQRQAYAKKLFKEALEELQITAKDLEHLNLIFPVSSSASSLL
                     VQLIREQWKESLGFAIPIVGKEFALLQADLSSGNFSLATGGWFADFADPMAFLTIFAY
                     PSGVPPYAINHKDFLEILQNIEQEQDHQKRSELVSQASLYLETFHIIEPIYHDAFQFA
                     MNKKLSNLGVSPTGVVDFRYAKEN"
BASE COUNT      560 a    439 c    294 g    506 t
ORIGIN


  Query Match           69.2%;  Score 25.6;  DB 6;  Length 1799;
  Best Local Similarity  87.5%;  Pred. No. 1.9;
  Matches   28;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      6 ggatcccatttttccttagcataacggaagtcc 37
          ||  |  ||||||||||||||||||||||||||
Db   1703 GGTGCTAATTTTCCTTAGCATAACGGAAGTCC 1672


RESULT   3
AE001606/c
LOCUS       AE001606               11648 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Chlamydia pneumoniae section 22 of 103 of the complete genome.
ACCESSION   AE001606 AE001363
VERSION     AE001606.1  GI:4376464
KEYWORDS    .
SOURCE      Chlamydophila pneumoniae CWL029.
  ORGANISM  Chlamydophila pneumoniae CWL029
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
REFERENCE   1  (bases 1 to 11648)
  AUTHORS   Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
            Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
  TITLE     Comparative genomes of Chlamydia pneumoniae and C. trachomatis
  JOURNAL   Nat. Genet. 21 (4), 385-389 (1999)
  MEDLINE   99206606
   PUBMED   10192388
REFERENCE   2  (bases 1 to 11648)
  AUTHORS   Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,
            Grimwood,J., Davis,R.W. and Stephens,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1998) Program in Infectious Diseases, University
            of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES             Location/Qualifiers
     source          1. .11648
                     /organism="Chlamydophila pneumoniae CWL029"
                     /strain="CWL029"
                     /db_xref="taxon:115713"
     gene            73. .1107
                     /gene="gcp_2"
                     /note="CPn0194"
     CDS             73. .1107
                     /gene="gcp_2"
                     /codon_start=1
                     /transl_table=11
                     /product="O-Sialoglycoprotein Endopeptidase"
                     /protein_id="AAD18347.1"
                     /db_xref="GI:4376465"
                     /translation="MLTLGLESSCDETACAIVNEDKQILANIIASQDIHASYGGVVPE
                     LASRAHLHIFPQVINKALQQANLLIEDMDLIAVTQTPGLIGSLSVGVHFGKGIAIGAK
                     KSLIGVNHVEAHLYAAYMAAQNVQFPALGLVVSGAHTAAFFIENPTSYKLIGKTRDDA
                     IGETFDKVGRFLGLPYPAGPLIEKLALEGSEDSYPFSPAKVPNYDFSFSGLKTAVLYA
                     IKGNNSSPRSPAPEISLEKQRDIAASFQKAACTTIAQKLPTIIKEFSCRSILIGGGVA
                     INEYFRSAIQTACNLPVYFPPAKLCSDNAAMIAGLGGENFQKNSSIPEIRICARYQWE
                     SVSPFSLASP"
     gene            1053. .2651
                     /gene="oppA_1"
                     /note="CPn0195"
     CDS             1053. .2651
                     /gene="oppA_1"
                     /codon_start=1
                     /transl_table=11
                     /product="Oligopeptide Binding Protein"
                     /protein_id="AAD18348.1"
                     /db_xref="GI:4376466"
                     /translation="MRKISVGICITILLSLSVVLQGCKESSHSSTSRGELAINIRDEP
                     RSLDPRQVRLLSEISLVKHIYEGLVQENNLSGNIEPALAEDYSLSSDGLTYTFKLKSA
                     FWSNGDPLTAEDFIESWKQVATQEVSGIYAFALNPIKNVRKIQEGHLSIDHFGVHSPN
                     ESTLVVTLESPTSHFLKLLALPVFFPVHKSQRTLQSKSLPIASGAFYPKNIKQKQWIK
                     LSKNPHYYNQSQVETKTITIHFIPDANTAAKLFNQGKLNWQGPPWGERIPQETLSNLQ
                     SKGHLHSFDVAGTSWLTFNINKFPLNNMKLREALASALDKEALVSTIFLGRAKTADHL
                     LPTNIHSYPEHQKQEMAQRQAYAKKLFKEALEELQITAKDLEHLNLIFPVSSSASSLL
                     VQLIREQWKESLGFAIPIVGKEFALLQADLSSGNFSLATGGWFADFADPMAFLTIFAY
                     PSGVPPYAINHKDFLEILQNIEQEQDHQKRSELVSQASLYLETFHIIEPIYHDAFQFA
                     MNKKLSNLGVSPTGVVDFRYAKEN"
     gene            2802. .4385
                     /gene="oppA_2"
                     /note="CPn0196"
     CDS             2802. .4385
                     /gene="oppA_2"
                     /codon_start=1
                     /transl_table=11
                     /product="Oligopeptide Binding Protein"
                     /protein_id="AAD18349.1"
                     /db_xref="GI:4376467"
                     /translation="MLRFFAVFISTLWLITSGCSPSQSSKGIFVVNMKEMPRSLDPGK
                     TRLIADQTLMRHLYEGLVEEHSQNGEIKPALAESYTISEDGTRYTFKIKNILWSNGDP
                     LTAQDFVSSWKEILKEDASSVYLYAFLPIKNARAIFDDTESPENLGVRALDKRHLEIQ
                     LETPCAHFLHFLTLPIFFPVHETLRNYSTSFEEMPITCGAFRPVSLEKGLRLHLEKNP
                     MYHNKSRVKLHKIIVQFISNANTAAILFKHKKLDWQGPPWGEPIPPEISASLHQDDQL
                     FSLPGASTTWLLFNIQKKPWNNAKLRKALSLAIDKDMLTKVVYQGLAEPTDHILHPRL
                     YPGTYPERKRQNERILEAQQLFEEALDELQMTREDLEKETLTFSTFSFSYGRICQMLR
                     EQWKKVLKFTIPIVGQEFFTIQKNFLEGNYSLTVNQWTAAFIDPMSYLMIFANPGGIS
                     PYHLQDSHFQTLLIKITQEHKKHLRNQLIIEALDYLEHCHILEPLCHPNLRIALNKNI
                     KNFNLFVRRTSDFRFIEKL"
     gene            4441. .5748
                     /gene="oppA_3"
                     /note="CPn0197"
```

```
CDS             4441. .5748
                /gene="oppA_3"
                /codon_start=1
                /transl_table=11
                /product="Oligopeptide Binding Protein"
                /protein_id="AAD18350.1"
                /db_xref="GI:4376468"
                /translation="MFSRWITLFLLFISLTGCSSYSSKHKQSLIIPIHDDPVAFSPEQ
                AKRAMDLSIAQLLFDGLTRETHRESNDLELAIASRYTVSEDFCSYTFFIKDSALWSDG
                TPITSEDIRNAWEYAQENSPHIQIFQGLNFSTPSSNAITIHLDSPNPDFPKLLAFPAF
                AIFKPENPKLFSGPYTLVEYFPGHNIHLKKNPNYYDYHCVSINSIKLLIIPDIYTAIH
                LLNRGKVDWVGQPWHQGIPWELHKQSQYHYYTYPVEGAFWLCLNTKSPHLNDLQNRHR
                LATCIDKRSIIEEALQGTQQPAETLSRGAPQPNQYKKQKPLTPQEKLVLTYPSDILRC
                QRIAEILKEQWKAAGIDLILEGLEYHLFVNKRKVQDYAIATQTGVAYYPGANLISEED
                KLLQNFEIIPIYYLSYDYLTQDFIEGVIYNASGAVDLKYTYFP"
gene            6032. .7612
                /gene="oppA_4"
                /note="CPn0198"
CDS             6032. .7612
                /gene="oppA_4"
                /codon_start=1
                /transl_table=11
                /product="Oligopeptide Binding Protein"
                /protein_id="AAD18351.1"
                /db_xref="GI:4376469"
                /translation="MHRLKPTLKSLIPNLLFLLLTLSSCSKQKQEPLGKHLVIAMSHD
                LADLDPRNAYLSRDASLAKALYEGLTRETDQGIALALAESYTLSKDHKVYTFKLRPSV
                WSDGTPLTAYDFEKSIKQLYFEEFSPSIHTLLGVIKNSSAIHNAQKSLETLGIQAKDD
                LTLVITLEQPFPYFLTLIARPVFSPVHHTLRESYKKGTPPSTYISNGPFVLKKHEHQN
                YLILEKNPHYYDHESVKLDRVTLKIIPDASTATKLFKSKSIDWIGSPWSAPISNEDQK
                VLSQEKILTYSVSSTTLLIYNLQKPLIQNKALRKAIAHAIDRKSILRLVPSGQEAVTL
                VPPNLSQLNLQKEISTEERQTKARAYFQEAKETLSEKELAELSILYPIDSSNSSIIAQ
                EIQRQLKDTLGLKIKIQGMEYHCFLKKRRQGDFFIATGGWIAEYVSPVAFLSILGNPR
                DLTQWRNSDYEKTLEKLYLPHAYKENLKRAEMIIEEETPIIPLYHGKYIYAIHPKIQN
                TFGSLLGHTDLKNIDILS"
gene            7905. .8849
                /gene="oppB_1"
                /note="CPn0199"
CDS             7905. .8849
                /gene="oppB_1"
                /codon_start=1
                /transl_table=11
                /product="Oligopeptide Permease"
                /protein_id="AAD18352.1"
                /db_xref="GI:4376470"
                /translation="MFSYIKNRILFNLLSLWIVLTLTFLVMKTIPGDPFNDEGCNVLS
                EEVLQTLKSRYGLDKPLYQQYTQYLHSIAKLDFGNSLVYKDRKVTNIISTAFPISAIL
                GLQSLFLSIGGGIALGTIAALKKKKQRRYILGASILQISIPAFIFATLLQYVFAVKIP
                LLPIACWGSFTHTILPTLALAVTPMAFIIQLTYSSVSAALNKDYVLLAYAKGLSPLKV
                VIKHILPYAIFPTISYSAFLTTTVITGTFAIENIFCIPGLGKWFICSIKQRDYPVALG
                LSVFYGTLFMLSSLLSDLIQSIIDPQIRYAHGKEKKRK"
gene            8880. .9734
                /gene="oppC_1"
                /note="CPn0200"
CDS             8880. .9734
                /gene="oppC_1"
                /codon_start=1
                /transl_table=11
                /product="Oligopeptide Permease"
                /protein_id="AAD18353.1"
                /db_xref="GI:4376471"
                /translation="MENLSSAPSRSIWKSIIQNKMLVLGLTTLIILMLGALLLPWFYQ
                DYEQTSLKDILVSPCSRFPFGTDTLGRCMFARTLRGLRLSLLIATIATLIDVCVGLLW
                ATVAISGGKKIDFLMMRTTEILFSLPRIPIIILLLVIFHHGLLPLILAMTITGWIPIS
                RIIYGQFLLLKNKPFVLSAKAMHASTFHILKKHLLPNTLAPIISTLIFTIPNAIYTEA
                FISFLGLGIQPPQASLGTLVKEGINAIDYYPWLFFFPSLIMIALSISFNLIGEGAKTL
                CLEEGSHG"
gene            9727. .10581
                /gene="oppD"
                /note="CPn0201"
CDS             9727. .10581
                /gene="oppD"
                /codon_start=1
                /transl_table=11
                /product="Oligopeptide Transport ATPase"
                /protein_id="AAD18354.1"
                /db_xref="GI:4376472"
                /translation="MDNYLLNIKDLTITSTNPKRTLIENLSLQLKENRNLALVGESGS
                GKTTITKAILGFLPENCLIKTGSILFEDIDITKLSPKELHKIRGQKIATILQNAMGSL
                TPSMRIGMQIIETLRQHHKMNKEEAYNKAMQLLTDVCIPNPKYSFSQYPFELSGGMRQ
                RVVIAIALASQPKLILADEPTTALDSMSQAQVLRILRNIQQQKQATILLVTHNLSLVK
                ELCNDICIIKDGKLIETGTVEEIFLSPKHPYTLKLLNAVSKIPIKKTSSPILKNKFQP
                LMSMQGGL"
gene            10578. .11366
                /gene="oppF"
                /note="CPn0202"
CDS             10578. .11366
                /gene="oppF"
                /codon_start=1
                /transl_table=11
                /product="Oligopeptide Transport ATPase"
                /protein_id="AAD18355.1"
                /db_xref="GI:4376473"
                /translation="MTTLLSIKDLSLTIRGKKILNHINLNLIKGSYLTIVGPSGSGKS
                SLALTILDLLKPTTGTITFHMDPKIPRARKVQVIWQDIDSSLNPCMSIKGIISEPLNI
                IGTYSKAEQNKEIYNVLDLVNLPKSVLHLKPYKLSGGQKQRIAIAKALVSKPELLICD
                EPLSSLDTLNQSLILDLFQTIKKEYQNTLLFITHDMSAAYYIADTIAVMDQGSLVEHA
                CREKIFSTPKHTTTQDLLDAIPIFSLISTEMEPSEEYELQVASK"
BASE COUNT      3743 a   2799 c   1831 g   3275 t
ORIGIN


  Query Match           69.2%;  Score 25.6;  DB 1;  Length 11648;
  Best Local Similarity  87.5%;  Pred. No. 2;
  Matches   28;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      6 ggatcccatttttccttagcataacggaagtcc 37
          ||  |  |||||||||||||||||||||||||||
Db   2655 GGTGCTAATTTTCCTTAGCATAACGGAAGTCC 2624


RESULT  4
AE002216
LOCUS       AE002216               11764 bp    DNA     linear   BCT 30-MAY-2000
DEFINITION  Chlamydophila pneumoniae AR39, section 44 of 94 of the complete
            genome.
ACCESSION   AE002216 AE002161
VERSION     AE002216.2  GI:8163460
KEYWORDS    .
SOURCE      Chlamydophila pneumoniae AR39.
  ORGANISM  Chlamydophila pneumoniae AR39
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
REFERENCE   1  (bases 1 to 11764)
  AUTHORS   Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
            White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
            Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
            Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
            McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
  TITLE     Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
            pneumoniae AR39
  JOURNAL   Nucleic Acids Res. 28 (6), 1397-1406 (2000)
  MEDLINE   20150255
   PUBMED   10684935
REFERENCE   2  (bases 1 to 11764)
  AUTHORS   Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
            White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
            Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
            Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
            McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Jun 1, 2000 this sequence version replaced gi:7189484.
FEATURES             Location/Qualifiers
     source          1. .11764
                     /organism="Chlamydophila pneumoniae AR39"
```

/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"

gene complement(109. .1707)
/gene="CP0572"

CDS complement(109. .1707)
/gene="CP0572"
/note="similar to GB:M57689 SP:P24141 GB:X56347 PID:143603 PID:40005; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, periplasmic peptide-binding protein, putative"
/protein_id="AAF38391.1"
/db_xref="GI:7189485"
/translation="MRKISVGICITILLSLSVVLQGCKESSHSSTSRGELAINIRDEP
RSLDPRQVRLLSEISLVKHIYEGLVQENNLSGNIEPALAEDYSLSSDGLTYTFKLKSA
FWSNGDPLTAEDFIESWKQVATQEVSGIYAFALNPIKNVRKIQEGHLSIDHFGVHSPN
ESTLVVTLESPTSHFLKLLALPVFFPVHKSQRTLQSKSLPIASGAFYPKNIKQKQWIK
LSKNPHYYNQSQVETKTITIHFIPDANTAAKLFNQGKLNWQGPPWGERIPQETLSNLQ
SKGHLHSFDVAGTSWLTFNINKFPLNNMKLREALASALDKEALVSTIFLGRAKTADHL
LPTNIHSYPEHQKQEMAQRQAYAKKLFKEALEELQITAKDLEHLNLIFPVSSSASSLL
VQLIREQWKESLGFAIPIVGKEFALLQADLSSGNFSLATGGWFADFADPMAFLTIFAY
PSGVPPYAINHKDFLEILQNIEQEQDHQKRSELVSQASLYLETFHIIEPIYHDAFQFA
MNKKLSNLGVSPTGVVDFRYAKEN"

gene complement(1653. .2687)
/gene="CP0573"

CDS complement(1653. .2687)
/gene="CP0573"
/note="O-sialoglycoprotein endopeptidase; identified by match to PFAM protein family HMM PF00814"
/codon_start=1
/transl_table=11
/product="O-sialoglycoprotein endopeptidase"
/protein_id="AAF73688.1"
/db_xref="GI:8163461"
/translation="MLTLGLESSCDETACAIVNEDKQILANIIASQDIHASYGGVVPE
LASRAHLHIFPQVINKALQQANLLIEDMDLIAVTQTPGLIGSLSVGVHFGKGIAIGAK
KSLIGVNHVEAHLYAAYMAAQNVQFPALGLVVSGAHTAAFFIENPTSYKLIGKTRDDA
IGETFDKVGRFLGLPYPAGPLIEKLALEGSEDSYPFSPAKVPNYDFSFSGLKTAVLYA
IKGNNSSPRSPAPEISLEKQRDIAASFQKAACTTIAQKLPTIIKEFSCRSILIGGGVA
INEYFRSAIQTACNLPVYFPPAKLCSDNAAMIAGLGGENFQKNSSIPEIRICARYQWE
SVSPFSLASP"

gene 2771. .3214
/gene="CP0574"

CDS 2771. .3214
/gene="CP0574"
/note="similar to GB:M17532 SP:P15282 GB:X13968 PID:145356 PID:43307; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="arginine repressor"
/protein_id="AAF38392.1"
/db_xref="GI:7189486"
/translation="MKKKVTIDEALKEILRLEGAATQEELCAKLLAQGFATTQSSVSR
WLRKIQAVKVAGERGARYSLPSSTEKTTTRHLVLSIRHNASLIVIRTVPGSASWIAAL
LDQGLKDEILGTLAGDDTIFVTPIDEGRLPLLMVSIANLLQVFLD"

gene 3263. .3916
/gene="CP0575"

CDS 3263. .3916
/gene="CP0575"
/note="similar to SP:P10345 PID:41572 GB:U00096 PID:1651370 PID:1787030; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="amino acid ABC transporter, permease protein"
/protein_id="AAF38393.1"
/db_xref="GI:7189487"
/translation="MDHWLAIARLLLRGCGYTLCVSGIGILCGSILGLLIGTVTSLYF
PSKLTKLLANSYVTVIRGTPLFIQILIIYFGLPEVLPIEPTPLVAGIIALSMNSAAYL
AENIRGGINSLSIGQWESAMVLGYKKYQIFVYIIYPQVFKNILPSLTNEFVSLIKESS
ILMVVGVPELTKVTKDIVSRELNPMEMYLICAGLYFLMTTSFSCISRLSEKRRSYDN"

gene 3906. .4586
/gene="CP0576"

CDS 3906. .4586
/gene="CP0576"
/note="similar to SP:P41076 PID:624632 GB:U00096 PID:1651271 PID:1778570; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="amino acid ABC transporter, ATP-binding protein"
/protein_id="AAF38394.1"
/db_xref="GI:7189488"
/translation="MTIRVRNLAYSVNKKKILDGVTFSLERGHITLFVGKSGSGKTMI
LRALAGLVQPTQGDIWIEGEAPALVFQQPELFSHMTVLGNCTHPQIHIKGRSTEEARE
KAFELLHLLDIEEVAKNYPDQLSGGQKQRVAIVRSLCMDKHTLLFDEPTSALDPFATA
SFRHLLETLRDQELTVGLTTHDMQFVHSCLDRIYLIDQGTVAGVYDKRDGELDSGHPL
SKYIHSAQ"

gene complement(4620. .5978)
/gene="CP0577"

CDS complement(4620. .5978)
/gene="CP0577"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38395.1"
/db_xref="GI:7189489"
/translation="MRKRHSFDSTSTKKEAVSKAIQKIIKIMETTDPSLNVETPNAEI
ESILQEIKEIKQKLSKQAEDLGLLEKYCSQETLSNLENTNASLKLSIGSVIEELASLK
QLVEESIEESLGQQDQLIQSVLIEISDKFLSSIGETLSGNLDMNQNVIQGLLIKENPE
KSEAASVGYVQTLLEPLSKRIGETHKKVATHDVNISSLQFHMMSVAGGRFRGHIDMNG
YRVLGLGEPKNGEDAVSKDYLERYVSSQLTIDKVEDKPITKPNKGKLLYSQGTSPKLE
GPLPLGLLTSGISGFTWKSASKSNDGSFPFSALRHKETESDTDCFQITSTTLSGNQAG
TYTWSLSLKVLVPSIFQIEKPEVQLSLVYSYEDWLPIDNIFNMSQPRTIPLALLGQTM
LAGQKYDILELAAHQTNQTLMISPNCSRFSLQLKQTNQFENSPVDFYIVHAAHSCHWS
GF"

gene complement(6069. .9497)
/gene="CP0578"

CDS complement(6069. .9497)
/gene="CP0578"
/note="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38396.1"
/db_xref="GI:7189490"
/translation="MKRRSWLKILGICLGSSIVLGFLIFLPQLLSTESGKYLVFSLIH
KESGLSCSAEELKISWFGRQTARKIKLTGEAKDEVFSAEKFELDGSLLRLLIYKKPKG
ITLSGWSLKINEPASIDHPSVSHLDPGSLLTYLNDCKIISEHGFITMKTVSGSSLSVS
GFYLEKSSEKFMTKCVVSEDQQSGNIFIESVLSPDVSISAQFSSVPVAFFKIFIASPF
WDHLLSYEDIINLSAEATHTNDGKISMTASGEGNQIQMKLQGHIHKSTFYIVEGSSSF
IELKPELASALCNQIIPLSTPITSKQIHATVSYAKIPLDITKWKHIEITSQAQLPEVA
IHPKDPNLALQLRDTKLGIKKTEKFSDIRYSSSTVLGGASPSHLNGLISIDNKKHLTK
FRLQQAQLPHTYLRAIFPQPFVINVPLDVAYYSLNIEGTYKNAHLEADAILDNPLLKL
SCSMSGAWKNFLFKGQGTYHFNKKWQEILSPHFSYAEARFSGKAQITDTNLFFPKFSG
KITARENELLIHAKFGSPNEPIKPETTSILIHGQFCSLPLSLVSNHLAPFHLKKLTFS
FHTDGGKFVTKGNLQALIENPDYPDLNNTRILIPDLLLSLDESSTSPSSKDLKIQGSG
EIFSLPLDSITKTYGKQVRLSPYFGSSGDLNFVVNYNPKDQNKLTLLSNFKSEALLGE
LKLVMDFSMKLSSGTQGTLQWEVSPERYASFFKNASCSPTCLLHRTANVRLDISKLSC
PEETKGLSCLTLLAAGGLEGSLEATPLIFYDNVSKETFIINDFKGSLRANNLDAKIEY
DLKGSCLAPRQDSKTLAEFSLEGQVDHLFSPESREFKQTANWIHIPSSFIAGIIPMSP
GLKAQISSLAGPRINVSIKNAFRFGEGPVDIMVDSENLQAQIPLILNEKSILLRENLT
AHLSINEDVNKAFLQEFNPLLAGGAYSQYPVTLEIDKQNFYLPIRPYSFEEFRIQSAT
LDMGKISIANTGTMYALFQFLDITDQKQFVESWFTPIFFSVQKGSIICKRXDALIDRR
IRLALWGKTDIAHDRLFMTLGIDPEVIKKYFHNTSLKTKNFFLIKIRGSISSPEVDWS
SAYARIALLKSYSLGNPFSSLADKLFSSLGDSTPPPTVHPFPWEKSNFDSIENK"

gene complement(9494. .10786)
/gene="CP0579"

CDS complement(9494. .10786)
/gene="CP0579"
/note="conserved hypothetical protein; identified by

```
                     Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF38397.1"
                     /db_xref="GI:7189491"
                     /translation="MFRKLFPFSKKKTGQKQRLRNNGLLQAIIQSIKVLLHNEAXKEA
                     CVLSYYGLLTCVPILVFFLRLSQHLFTNLNWKEWLIIKFPDYKKPIVAIVEAAYHATE
                     SNIGLVLVGSFFVFCWAGILMLLSLEDGLNKIFRTSWTPISLKRLVSYFVITLVSPMI
                     FIIVCGSWIYITQIMPIQYAKLFSLSHSMTALYFISRFVPYLLLYLALFCCYAFLPRV
                     AIQKTSALISTLIIGSVWIVFQKAFFSLQVSIFNYSFTYGALVALPSFLLLLYIYTMI
                     YLFGGALTFIIQNRGCTFIFLGDKILPSCYLQLITSTYILALTTRQFNEGLSPLTAQF
                     IAKQSKVPIGEVSQCLDVLEKEGFLFPYNNGYQPVFNFSELTIKDIADKLLHREIFKK
                     FNPDLGITFIENSFQNIFNQASKNKENLTLSEIARRIK"
     gene            complement(10849. .11649)
                     /gene="CP0580"
     CDS             complement(10849. .11649)
                     /gene="CP0580"
                     /note="conserved hypothetical protein; identified by
                     Glimmer2; putative"
                     /codon_start=1

  Query Match             69.2%;  Score 25.6;  DB 1;  Length 11764;
  Best Local Similarity   87.5%;  Pred. No. 2;
  Matches   28;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       6 ggatcccatttttccttagcataacggaagtcc 37
           ||  |  ||||||||||||||||||||||||||
Db     105 GGTGCTAATTTTCCTTAGCATAACGGAAGTCC 136


RESULT   5
AP002545/c
LOCUS       AP002545               299650 bp    DNA     linear   BCT 08-DEC-2000
DEFINITION  Chlamydophila pneumoniae J138 genomic DNA, complete sequence,
            section 1/4.
ACCESSION   AP002545 AB033780 AB033781 AB033792 AB033793 AB033794 AB033795
            AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073
            AB036074 AB036075 AB036076 AB036077 AB036078 AB038345 AB038346
            AB038347 BA000008
VERSION     AP002545.2  GI:9956082
KEYWORDS    .
SOURCE      Chlamydophila pneumoniae J138 (strain:J138) DNA.
  ORGANISM  Chlamydophila pneumoniae J138
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
REFERENCE   1  (sites)
  AUTHORS   Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
            Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.
  TITLE     Comparison of whole genome sequences of Chlamydia pneumoniae J138
            from Japan and CWL029 from USA
  JOURNAL   Nucleic Acids Res. 28 (12), 2311-2314 (2000)
  MEDLINE   20330349
REFERENCE   2  (bases 1 to 299650)
  AUTHORS   Shirai,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University
            School of Medicine, Department of Microbiology; 1-1-1
            Minamikogushi, Ube, Yamaguchi 755-8505, Japan
            (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227,
            Fax:81-836-22-2415)
COMMENT     On Aug 31, 2000 this sequence version replaced gi:6172286
            gi:6172288 gi:6172310 gi:6172312 gi:6172314 gi:6172316 gi:6172318
            gi:6172320 gi:6172322 gi:6172324 gi:6635158 gi:6635160 gi:6635162
            gi:6635164 gi:6635166 gi:6635168 gi:6635170 gi:6635172 gi:8547422
            gi:8547426 gi:8547430 gi:8978373.
            AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)
            AB038345-AB038347: Submitted (14-Feb-2000)
            AB036071-AB036078: Submitted (18-Dec-2000).
FEATURES              Location/Qualifiers
     source           1. .299650
                      /organism="Chlamydophila pneumoniae J138"
                      /strain="J138"
                      /db_xref="taxon:138677"
                      /note="synonym:Chlamydia pneumoniae (strain J138)"
     gene             complement(1. .282)
                      /gene="CPj0001"
     CDS              complement(1. .282)
                      /gene="CPj0001"
                      /codon_start=1
                      /transl_table=11
                      /product="CT001 hypothetical protein"
                      /protein_id="BAA98211.1"
                      /db_xref="GI:8978374"
                      /translation="MLGKIIRGLSSLIVILCALNVGLIGITHNKLNIIAKLCGGVSTP
                      ATQITYIIIGIAGVICLLSFCPFCSKKSRHSHGDSCSSGGCHSHHSDKN"
     gene             573. .878
                      /gene="gatC"
     CDS              573. .878
                      /gene="gatC"
                      /codon_start=1
                      /transl_table=11
                      /product="Glu-tRNA Gln amidotransferase (C subunit)"
                      /protein_id="BAA98212.1"
                      /db_xref="GI:8978375"
                      /translation="MEQFHLDREEILLLAKASALQLSEELIQEVQTSLSAVITSMKEA
                      LAIEIDDADSCESLFMHVVNVEDLREDSVTSDFNREEFLRNVPESLGGLVKVPAVIK"
     gene             895. .2373
                      /gene="gatA"
     CDS              895. .2373
                      /gene="gatA"
                      /codon_start=1
                      /transl_table=11
                      /product="Glu tRNA Gln amidotransferae (A subunit)"
                      /protein_id="BAA98213.1"
                      /db_xref="GI:8978376"
                      /translation="MYRYSALELAKAVTLGELTATGVTQHFFHRIEEAEGQVGAFISL
                      CKEQALEQAELIDKKRSRGEPLGKLAGVPVGIKDNIHVTGLKTTCASRVLENYQPPFD
                      ATVVERIKKEDGIILGKLNMDEFAMGSTTLYSAFHPTHNPWDLSRVPGGSSGGSAAAV
                      SARFCPVALGSDTGGSIRQPAAFCGVVGFKPSYGAVSRYGLVAFASSLDQIGPLANTV
                      EDVALMMDVFSGRDPKDATSREFFRDSFMSKLSTEVPKVIGVPRTFLEGLRDDIRENF
                      FSSLAIFEGEGTHLVDVELDILSHAVSIYYILASAEAATNLARFDGVRYGYRSPQAHT
                      ISQLYDLSRGEGFGKEVMRRILLGNYVLSAERQNVYYKKATAVRAKIVKAFRTAFEKC
                      EILAMPVCSSPAFEIGEILDPVTLYLQDIYTVAMNLAYLPAIAVPSGFSKEGLPLGLQ
                      IIGQQGQDQQVCQVGYSFQEHAQ1KQLFSKRYAKSVVLGGQS"
     gene             2370. .3836
                      /gene="gatB"
     CDS              2370. .3836
                      /gene="gatB"
                      /codon_start=1
                      /transl_table=11
                      /product="(Pet112) Glu tRNA Gln amidotransferase (B
                      subunit)"
                      /protein_id="BAA98214.1"
                      /db_xref="GI:8978377"
                      /translation="MSAVYADWESVIGLEVHVELNTASKLFSSALNRFGDEPNTNIST
                      VCTGLPGSLPVLNQSAVEKAVLFGCAVEGEISLLSRFDRKSYFYPDSPRNFQITQFEH
                      PIIRGGRIKAIVQGEERYFELAQTHIEDDAGMLKHFGEFAGVDYNRAGVPLIEIVSKP
                      CMFCPEDAVAYATSLVSLLDYIGISDCNMEEGSIRFDVNVSVRPKGSPELRNKVEIKN
                      MNSFAFMAQALEAEKQRQIDEYLNQPNKDPKLVIPAATYRWDPEKKKTVLMRLKESAE
                      DYKYFPEPDLPTLQLTESYIERIRKTLPELPYDKYHRYIQEYGLSEDIASILISDKNI
                      ATFFEVACKDCKNFRSLSNWVSVEFGGRCKTLGVKLPSSGIFPEGVAQLVNAIDQGVI
                      TGKIAKEIADLMMESPGKNPEEILKEKPELLPMSDEGELQKIIAEVVLANPESIVDYK
                      NGKTKALGFLVGQIMKRTAGKAPPKRVNELLLLELDKG"
     gene             4127. .6895
                      /gene="pmp_1"
     CDS              4127. .6895
                      /gene="pmp_1"
                      /codon_start=1
                      /transl_table=11
                      /product="polymorphic outer membrane protein G family"
                      /protein_id="BAA98215.1"
                      /db_xref="GI:8978378"
                      /translation="MRFSLCGFPLVFSFTLLSVFDTSLSATTISLTPEDSFHGDSQNA
                      ERSYNVQAGDVYSLTGDVSISNVDNSALNKACFNVTSGSVTFAGNHHGLYFNNISSGT
                      TKEGAVLCCQDPQATARFSGFSTLSFIQSPGDIKEQGCLYSKNALMLLNNYVVRFEQN
                      QSKTKGGAISGANVTIVGNYDSVSFYQNAATFGGAIHSSGPLQIAVNQAEIRFAQNTA
```

```
                KNGSGGALYSDGDIDIDQNAYVLFRENEALTTAIGKGGAVCCLPTSGSSTPVPIVTFS
                DNKQLVFERNHSIMGGGAIYARKLSISSGGPTLFINNISYANSQNLGGAIAIDTGGEI
                SLSAEKGTITFQGNRTSLPFLNGIHLLQNAKFLKLQARNGYSIEFYDPITSEADGSTQ
                LNINGDPKNKEYTGTILFSGEKSLANDPRDFKSTIPQNVNLSAGYLVIKEGAEVTVSK
                FTQSPGSHLVLDLGTKLIASKEDIAITGLAIDIDSLSSSSTAAVIKANTANKQISVTD
                SIELISPTGNAYEDLRMRNSQTFPLLSLEPGAGGSVTVTAGDFLPVSPHYGFQGNWKL
                AWTGTGNKVGEFFWDKINYKPRPEKEGNLVPNILWGNAVDVRSLMQVQETHASSLQTD
                RGLWIDGIGNFFHVSASEDNIRYRHNSGGYVLSVNNEITPKHYTSMAFSQLFSRDKDY
                AVSNNEYRMYLGSYLYQYTTSLGNIFRYASRNPNVNVGILSRRFLQNPLMIFHFLCAY
                GHATNDMKTDYANFPMVKNSWRNNCWAIECGGSMPLLVPENGRLFQGAIPFMKLQLVY
                AYQGDFKETTADGRRFSNGSLTSISVPLGIRFEKLALSQDVLYDFSFSYIPDIFRKDP
                SCEAALVISGDSWLVPAAHVSRHAFVGSGTGRYHFNDYTELLCRGSIECRPHARNYNI
                NCGSKFRF"
gene            complement(7138. .7293)
                /gene="CPj0006"
CDS             complement(7138. .7293)
                /gene="CPj0006"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAA98216.1"
                /db_xref="GI:8978379"
                /translation="MQEPLRSALLERLSEWLVLLGVPSPETTRSTPEKDANQLPKDSR
                NRTLESL"
gene            7515. .10169
                /gene="CPj0007"
CDS             7515. .10169
                /gene="CPj0007"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAA98217.2"
                /db_xref="GI:9956083"
                /translation="MFSFLVVIPLTDSTTSSLSTSLLDEGNPQSMRKLRILAIVLIAL
                SIILIAGGVVLLTVAIPGLSSVISSPAGMGACALGCVMLALGIDVLLKKREVPIVLAS
                VTTTPGTVSPRSGISISGADSTIRSLPTYLLDEGHPQSMRKLRILAIVLIVFSIILIA
                SGVVLLTVAIPGLSSIISSPAEMGACALGCVMLALGIDVLLKKREVPIVVPAPIPEEV
                VIDDIDEESIRLQQEAEAALARLPEEMSAFEGYIKVVESHLENMKSLPYDGHGLEEKT
                KHQIRVVRSSLKAMVPEFLDIRRIFEEEEFFFLSARKRLIDLATTLVERKILTEQLER
                NNLRKAFSYLYQDSIFKKIIDNFEKLAWKFMILSKSICRFTIIFENHEHGVAKSLLHK
                NAVLLEKVIYRSLQKSYRDIGMSSAKMKILHGNPFFSLEDNKKTIMKEHAEMLESLSS
                YRKVFLALSDENVVDTPSDPKKWDLSGIPCRDALSEISRDEQWQKKAHLKHQESLYTQ
                ARDRLTDQSSKENQKELEKAEQEYISSWERVKKFEIERVQERIQAIQKLYPNILEREE
                ETTGQETVTPTVQGTTASSDLTDILGRIEVSSREDNQNQESCVKVLRSHEVEMSWEVK
                QEYGPKKKEFQDQMGSLERFFTEHIEELEVLQKDYSKHLSYFKKVNNKKEVQYAKFRL
                KVLESDLEGILAQTESAESLLTQEELPILATRGALEKAVFKGSLCCALASKAKPYFEE
                DPRFQDSDTQLRALTLRLQEAKASLEEEIKRFSNLENDIAEERRLLKESKQTFERAGL
                GVLREIAVESTYDLRSLTNTWEGTPESEKVYFSMYLNYYNEEKRRAKTRLVEMTQRYR
                DFKMALEAMQFNEEALLQEELSIQAPSE"
gene            10645. .11358
                /gene="CPj0008"
CDS             10645. .11358
                /gene="CPj0008"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAA98218.1"
                /db_xref="GI:8978381"
                /translation="MISGLLFLLVRREVPTVRSEEIPRGVSVTPSEEPALEKAQKEPE
                TKKILDRLPKELDQLDTYIQEVFACLERLKDPKYEDRGLLTEAKEKLRVFDVVEKDMM
                SEFLDIQRVLNEEAYYVEHCQDPLENIAYEIFSSQELRDYYCAGVCGYLPSGDARADR
                LKRSVKEVMDRFMRVTWKSWEASVMLDHSYGVARELFKKAVGVLEESVYKILFKSYRD
                AFYECEKAKIQRDGRFKWL"
gene            11485. .12792
                /gene="CPj0009"
CDS             11485. .12792
                /gene="CPj0009"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAA98219.1"
                /db_xref="GI:8978382"
                /translation="MWLDRYADKFILREKEEKMERHELFHATMVRKASGHAYAKAKAA
                FEKERSNENQRKVKDVEKWLSKGLAEFRNQESRRARERLRELQTLYPEVSVEERVLER
                QRTKKVNLENLYADIEKKYHHCVREQEHYWKEVENKEAEYRENGEKVLSAEEVSECLQ
                RLEDCLETWSKKLTKAEESVFEMKFDATEKLGNKVLSDVTNRLEILCEDAEEMIFRIE
                EIEMTLRMVELPLLFMKNTFEKASLQYNSCKEMLAKVEPQCKESPTYRSSQERLERLN
                QDLQTAYTNCQERLQGFSDLESKVRTCRDHLREQMKHFEVQGLNFINEELLWVGAELF
                TQARLDLVATVPYMEFYLQYHNIKREKVRSQWMAKTERYREIRQAFQGVMKEDLLAED
                TILKEEDYWLLRDDWLLRDERKNRQRRLICNKIAAAQQRVKGGF"
gene            13104. .15428
                /gene="CPj0010"
CDS             13104. .15428
                /gene="CPj0010"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"

  Query Match             69.2%;  Score 25.6;  DB 1;  Length 299650;
  Best Local Similarity   87.5%;  Pred. No. 2.4;
  Matches   28;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       6 ggatcccattttccttagcataacggaagtcc 37
           ||  |  ||||||||||||||||||||||||||
Db  235407 GGTGCTAATTTTCCTTAGCATAACGGAAGTCC 235376


RESULT   6
AX349501/c
LOCUS       AX349501               1599 bp    DNA     linear   PAT 06-FEB-2002
DEFINITION  Sequence 24 from Patent WO0202606.
ACCESSION   AX349501
VERSION     AX349501.1  GI:18615357
KEYWORDS    .
SOURCE      Chlamydophila pneumoniae.
  ORGANISM  Chlamydophila pneumoniae
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
REFERENCE   1  (sites)
  AUTHORS   Ratti,G. and Grandi,G.
  TITLE     Immunisation against Chlamydia pneumoniae
  JOURNAL   Patent: WO 0202606-A 24 10-JAN-2002;
            Chiron S.p.A. (IT)
FEATURES             Location/Qualifiers
     source          1. .1599
                     /organism="Chlamydophila pneumoniae"
                     /db_xref="taxon:83558"
BASE COUNT      509 a    389 c    261 g    440 t
ORIGIN


  Query Match             67.6%;  Score 25;  DB 6;  Length 1599;
  Best Local Similarity   100.0%;  Pred. No. 3.4;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 attttccttagcataacggaagtcc 37
           |||||||||||||||||||||||||
Db    1596 ATTTTCCTTAGCATAACGGAAGTCC 1572


RESULT   7
AL591675/c
LOCUS       AL591675               166919 bp    DNA     linear   HTG 06-JUN-2001
DEFINITION  Mus musculus chromosome 2 clone RP23-420L2, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL591675
VERSION     AL591675.5  GI:14330189
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 166919)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
```

```
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 8, 2001 this sequence version replaced gi:14270837.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM420L2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 165475 bases at least Q40
            Consensus quality: 165847 bases at least Q30
            Consensus quality: 166059 bases at least Q20
            Insert size: 166219; sum-of-contigs
            Insert size: 167305; 4.6% error; agarose-fp
            Quality coverage: 10.11x in Q20 bases; sum-of-contigs Quality
            coverage: 10.17x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES                Location/Qualifiers
     source             1. .166919
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /chromosome="2"
                        /clone="RP23-420L2"
                        /clone_lib="RPCI-23"
     misc_feature       1. .15392
                        /note="assembly_fragment:01294
                        clone_end:SP6
                        vector_side:left"
     misc_feature       15493. .33097
                        /note="assembly_fragment:00072
                        fragment_chain:1"
     misc_feature       33198. .53290
                        /note="assembly_fragment:04398
                        fragment_chain:1"
     misc_feature       53391. .71851
                        /note="assembly_fragment:03435
                        fragment_chain:1"
     misc_feature       71952. .82212
                        /note="assembly_fragment:04350
                        fragment_chain:1"
     misc_feature       82313. .93447
                        /note="assembly_fragment:00274
                        fragment_chain:2"
     misc_feature       93548. .141641
                        /note="assembly_fragment:01130
                        fragment_chain:2"
     misc_feature       141742. .166919
                        /note="assembly_fragment:04396
                        fragment_chain:2
                        clone_end:T7
                        vector_side:right"
BASE COUNT    48574 a  33038 c  33509 g  51098 t    700 others
ORIGIN


  Query Match             58.4%;  Score 21.6;  DB 2;  Length 166919;
  Best Local Similarity   75.0%;  Pred. No. 1.3e+02;
  Matches   27;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 gcgccggatcccatttccttagcataacggaagtc 36
           | |||  |||| | || |||| || ||| |||||||
Db   78647 GTGCCAAATCCTACTTCCCTTTGCCTAATGGAAGTC 78612


RESULT   8
```

```
AL626766
LOCUS       AL626766              187334 bp    DNA     linear   HTG 21-OCT-2001
DEFINITION  Mus musculus chromosome 2 clone RP23-345I2, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL626766
VERSION     AL626766.7  GI:16304551
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 187334)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 21, 2001 this sequence version replaced gi:16151503.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM345I2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 184803 bases at least Q40
            Consensus quality: 185238 bases at least Q30
            Consensus quality: 185716 bases at least Q20
            Insert size: 186434; sum-of-contigs
            Insert size: 181299; 2.7% error; agarose-fp
            Quality coverage: 11.78x in Q20 bases; sum-of-contigs Quality
            coverage: 12.11x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES                Location/Qualifiers
     source             1. .187334
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /chromosome="2"
                        /clone="RP23-345I2"
                        /clone_lib="RPCI-23"
     misc_feature       1. .4771
                        /note="assembly_fragment:04328
                        fragment_chain:1
                        clone_end:SP6
                        vector_side:left"
     misc_feature       4872. .16238
                        /note="assembly_fragment:04945
                        fragment_chain:1"
     misc_feature       16339. .18353
                        /note="assembly_fragment:04074
                        fragment_chain:1"
     misc_feature       18454. .29075
                        /note="assembly_fragment:05173
                        fragment_chain:1"
     misc_feature       29176. .47284
                        /note="assembly_fragment:04638
                        fragment_chain:2"
     misc_feature       47385. .140964
                        /note="assembly_fragment:04370
                        fragment_chain:2"
     misc_feature       141065. .144487
                        /note="assembly_fragment:03914
                        fragment_chain:2"
     misc_feature       144588. .154511
                        /note="assembly_fragment:02762
```

```
                     fragment_chain:2"
     misc_feature    154612. .180806
                     /note="assembly_fragment:05090
                     fragment_chain:2"
     misc_feature    180907. .187334
                     /note="assembly_fragment:01106
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
BASE COUNT    54613 a  39041 c  37442 g  55325 t    913 others
ORIGIN


  Query Match           58.4%;  Score 21.6;  DB 2;  Length 187334;
  Best Local Similarity  75.0%;  Pred. No. 1.3e+02;
  Matches   27;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 gcgccggatcccatttttccttagcataacggaagtc 36
          | |||  |||| | || |||| || ||| |||||||
Db  22063 GTGCCAAATCCTACTTCCCTTTGCCTAATGGAAGTC 22098


RESULT   9
AC105488/c
LOCUS       AC105488               226999 bp    DNA     linear   HTG 09-JAN-2002
DEFINITION  Rattus norvegicus chromosome Rf1 clone CH230-104J23, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC105488
VERSION     AC105488.1  GI:18092710
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 226999)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
    TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 226999)
  AUTHORS   Worley,K.C.
    TITLE   Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GNBA
                Center clone name: CH230-104J23
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 215114 bases at least Q40
                Consensus quality: 216690 bases at least Q30
                Consensus quality: 217874 bases at least Q20
                Estimated insert size: 219208; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    71450: contig of 71450 bp in length
            *    71451    71550: gap of unknown length
            *    71551   107586: contig of 36036 bp in length
            *   107587   107686: gap of unknown length
            *   107687   140791: contig of 33105 bp in length
            *   140792   140891: gap of unknown length
            *   140892   155843: contig of 14952 bp in length
            *   155844   155943: gap of unknown length
            *   155944   170763: contig of 14820 bp in length
            *   170764   170863: gap of unknown length
            *   170864   180907: contig of 10044 bp in length
            *   180908   181007: gap of unknown length
            *   181008   194123: contig of 13116 bp in length
            *   194124   194223: gap of unknown length
            *   194224   204722: contig of 10499 bp in length
            *   204723   204822: gap of unknown length
            *   204823   214668: contig of 9846 bp in length
            *   214669   214768: gap of unknown length
            *   214769   219638: contig of 4870 bp in length
            *   219639   219738: gap of unknown length
            *   219739   223451: contig of 3713 bp in length
            *   223452   223551: gap of unknown length
            *   223552   224579: contig of 1028 bp in length
            *   224580   224679: gap of unknown length
            *   224680   226999: contig of 2320 bp in length.
FEATURES             Location/Qualifiers
     source          1. .226999
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /chromosome="Rf1"
                     /clone="CH230-104J23"
BASE COUNT    58605 a  51250 c  53679 g  62221 t   1244 others
ORIGIN
```

```
  Query Match             58.4%;  Score 21.6;  DB 2;  Length 226999;
  Best Local Similarity   75.0%;  Pred. No. 1.3e+02;
  Matches   27;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      2 cgccggatcccatttccttagcataacggaagtcc 37
          ||| ||  ||  |||||||| || | || |||||||
Db 156345 CGCTGGTGCCTGTTTTCCTTTGCTTGACAGAAGTCC 156310


RESULT 10
AC097752/c
LOCUS       AC097752              239486 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus chromosome Rf1 clone CH230-75H6, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC097752
VERSION     AC097752.2  GI:17944091
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 239486)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 239486)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:16327457.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFAW
                Center clone name: CH230-75H6
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 220015 bases at least Q40
                Consensus quality: 225080 bases at least Q30
                Consensus quality: 228942 bases at least Q20
                Estimated insert size: 228246; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    27619: contig of 27619 bp in length
            *    27620    27719: gap of unknown length
            *    27720    45331: contig of 17612 bp in length
            *    45332    45431: gap of unknown length
            *    45432    62033: contig of 16602 bp in length
            *    62034    62133: gap of unknown length
            *    62134    73808: contig of 11675 bp in length
            *    73809    73908: gap of unknown length
            *    73909    94892: contig of 20984 bp in length
            *    94893    94992: gap of unknown length
            *    94993   111560: contig of 16568 bp in length
            *   111561   111660: gap of unknown length
            *   111661   124503: contig of 12843 bp in length
            *   124504   124603: gap of unknown length
            *   124604   134076: contig of 9473 bp in length
            *   134077   134176: gap of unknown length
            *   134177   141809: contig of 7633 bp in length
            *   141810   141909: gap of unknown length
            *   141910   147067: contig of 5158 bp in length
            *   147068   147167: gap of unknown length
            *   147168   153390: contig of 6223 bp in length
            *   153391   153490: gap of unknown length
            *   153491   158165: contig of 4675 bp in length
            *   158166   158265: gap of unknown length
            *   158266   163900: contig of 5635 bp in length
            *   163901   164000: gap of unknown length
            *   164001   168008: contig of 4008 bp in length
            *   168009   168108: gap of unknown length
            *   168109   172574: contig of 4466 bp in length
            *   172575   172674: gap of unknown length
            *   172675   178530: contig of 5856 bp in length
            *   178531   178630: gap of unknown length
            *   178631   182308: contig of 3678 bp in length
            *   182309   182408: gap of unknown length
            *   182409   187521: contig of 5113 bp in length
            *   187522   187621: gap of unknown length
            *   187622   193073: contig of 5452 bp in length
            *   193074   193173: gap of unknown length
            *   193174   197753: contig of 4580 bp in length
            *   197754   197853: gap of unknown length
            *   197854   202763: contig of 4910 bp in length
            *   202764   202863: gap of unknown length
            *   202864   207689: contig of 4826 bp in length
            *   207690   207789: gap of unknown length
            *   207790   210513: contig of 2724 bp in length
            *   210514   210613: gap of unknown length
            *   210614   214864: contig of 4251 bp in length
```

```
            *   214865   214964: gap of unknown length
            *   214965   218656: contig of 3692 bp in length
            *   218657   218756: gap of unknown length
            *   218757   221049: contig of 2293 bp in length
            *   221050   221149: gap of unknown length
            *   221150   223336: contig of 2187 bp in length
            *   223337   223436: gap of unknown length
            *   223437   224923: contig of 1487 bp in length
            *   224924   225023: gap of unknown length
            *   225024   227491: contig of 2468 bp in length
            *   227492   227591: gap of unknown length
            *   227592   228771: contig of 1180 bp in length
            *   228772   228871: gap of unknown length
            *   228872   230889: contig of 2018 bp in length
            *   230890   230989: gap of unknown length
            *   230990   232730: contig of 1741 bp in length
            *   232731   232830: gap of unknown length
            *   232831   234272: contig of 1442 bp in length
            *   234273   234372: gap of unknown length
            *   234373   235604: contig of 1232 bp in length
            *   235605   235704: gap of unknown length
            *   235705   237074: contig of 1370 bp in length
            *   237075   237174: gap of unknown length
            *   237175   238322: contig of 1148 bp in length
            *   238323   238422: gap of unknown length
            *   238423   239486: contig of 1064 bp in length.
FEATURES             Location/Qualifiers
     source          1. .239486
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /chromosome="Rf1"
                     /clone="CH230-75H6"
BASE COUNT    64295 a  53502 c  53422 g  64642 t   3625 others
ORIGIN


  Query Match           58.4%;  Score 21.6;  DB 2;  Length 239486;
  Best Local Similarity 75.0%;  Pred. No. 1.3e+02;
  Matches   27;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       2 cgccggatcccatttttccttagcataacggaagtcc 37
           ||| ||  ||  |||||||| || | || |||||||
Db 162498 CGCTGGTGCCTGTTTTCCTTTGCTTGACAGAAGTCC 162463


RESULT  11
AC069483
LOCUS       AC069483               151321 bp    DNA     linear   PRI 25-NOV-2001
DEFINITION  Homo sapiens, clone RP11-29M5, complete sequence.
ACCESSION   AC069483
VERSION     AC069483.6  GI:17064696
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151321)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-29M5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151321)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 151321)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 25, 2001 this sequence version replaced gi:16756289.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10441
                Center clone name: 29_M_5
            --------------.
FEATURES             Location/Qualifiers
     source          1. .151321
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-29M5"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   920. .1158
                     /rpt_family="MIR"
     repeat_region   1519. .1541
                     /rpt_family="AT_rich"
     repeat_region   1829. .2044
                     /rpt_family="AluJb"
     repeat_region   2057. .2351
                     /rpt_family="AluJo"
     repeat_region   complement(2608. .2684)
                     /rpt_family="MIR"
     repeat_region   complement(5550. .5748)
                     /rpt_family="MIR"
     repeat_region   5809. .5935
                     /rpt_family="FLAM_A"
     repeat_region   7165. .7216
                     /rpt_family="AT_rich"
```

```
repeat_region   8907. .9195
                /rpt_family="AluJo"
repeat_region   complement(9196. .9478)
                /rpt_family="AluSx"
repeat_region   9488. .9510
                /rpt_family="AT_rich"
repeat_region   9768. .9789
                /rpt_family="AT_rich"
repeat_region   11461. .11483
                /rpt_family="(TTTA)n"
repeat_region   complement(13745. .13867)
                /rpt_family="L2"
repeat_region   14034. .14220
                /rpt_family="MIR"
repeat_region   14417. .14722
                /rpt_family="AluJb"
repeat_region   complement(14991. .15055)
                /rpt_family="MER116"
repeat_region   16645. .16866
                /rpt_family="MIR3"
repeat_region   complement(17501. .17900)
                /rpt_family="L1ME1"
repeat_region   complement(18075. .18202)
                /rpt_family="MIR"
repeat_region   18203. .18303
                /rpt_family="HY3"
repeat_region   19347. .19655
                /rpt_family="AluSx"
repeat_region   complement(19674. .19845)
                /rpt_family="L1MB7"
repeat_region   22277. .22296
                /rpt_family="(TA)n"
repeat_region   23628. .23654
                /rpt_family="(TTC)n"
repeat_region   complement(24237. .24594)
                /rpt_family="L2"
repeat_region   24962. .25061
                /rpt_family="MIR"
repeat_region   25148. .25452
                /rpt_family="AluSc"
repeat_region   26584. .26712
                /rpt_family="MIR"
repeat_region   complement(26967. .27238)
                /rpt_family="AluJo"
repeat_region   complement(27810. .27952)
                /rpt_family="MER5A"
repeat_region   complement(28204. .28348)
                /rpt_family="MER5A"
repeat_region   29028. .29062
                /rpt_family="(CA)n"
repeat_region   complement(34139. .34273)
                /rpt_family="MIR3"
repeat_region   36575. .36597
                /rpt_family="(CAAA)n"
repeat_region   37037. .37270
                /rpt_family="AluJo"
repeat_region   37273. .37317
                /rpt_family="(CA)n"
repeat_region   37685. .37711
                /rpt_family="(CA)n"
repeat_region   37718. .37765
                /rpt_family="GA-rich"
repeat_region   37876. .38051
                /rpt_family="MIR"
repeat_region   38194. .38492
                /rpt_family="AluJo"
repeat_region   38497. .38753
                /rpt_family="AluSx"
repeat_region   complement(39131. .39251)
                /rpt_family="FLAM_C"
repeat_region   39305. .39411
                /rpt_family="MIR"
repeat_region   41050. .41368
                /rpt_family="L1MC1"
repeat_region   43352. .43379
                /rpt_family="AT_rich"
repeat_region   43720. .43839
                /rpt_family="MIR"
repeat_region   44936. .45221
                /rpt_family="AluSq"
repeat_region   45625. .45873
                /rpt_family="L1ME4A"
repeat_region   45963. .46178
                /rpt_family="L1ME4A"
repeat_region   46179. .46490
                /rpt_family="AluSg"
repeat_region   46491. .46583
                /rpt_family="L1ME4A"
repeat_region   46585. .46783
                /rpt_family="AluJo"
repeat_region   46903. .47289
                /rpt_family="L1ME4A"
repeat_region   47404. .47427
                /rpt_family="AT_rich"
repeat_region   complement(47968. .48272)
                /rpt_family="MLT1B"
repeat_region   49874. .49970
                /rpt_family="MIR"
repeat_region   50992. .51146
                /rpt_family="MIR"
repeat_region   complement(51519. .51809)

  Query Match           57.8%;  Score 21.4;  DB 9;  Length 151321;
  Best Local Similarity  80.6%;  Pred. No. 1.6e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      6 ggatcccatttttccttagcataacggaagtc 36
          || |||| |||| |||| |||||| |||| |
Db  30449 GGGTCCCTTTTTTCTTATCATAACAGAAGGC 30479


RESULT  12
AC078909
LOCUS       AC078909              170257 bp    DNA     linear   PRI 12-DEC-2001
DEFINITION  Homo sapiens chromosome 15, clone RP11-128A17, complete sequence.
ACCESSION   AC078909
VERSION     AC078909.7  GI:17530779
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170257)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-128A17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170257)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
```

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 170257)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 170257)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Dec 12, 2001 this sequence version replaced gi:15799631.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L9136
Center clone name: 128_A_17
----------.

FEATURES Location/Qualifiers
source 1. .170257
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-128A17"
/clone_lib="RPCI-11 Human Male BAC"
repeat_region complement(1. .99)
/rpt_family="MER1B"
repeat_region 209. .394
/rpt_family="AluSc"
repeat_region 1020. .1068
/rpt_family="(TC)n"
repeat_region 1068. .1104
/rpt_family="(TG)n"
repeat_region 2513. .2533
/rpt_family="AT_rich"
repeat_region complement(2616. .2711)
/rpt_family="L2"
repeat_region complement(2918. .3086)
/rpt_family="MIR"
repeat_region 4316. .4347
/rpt_family="(TA)n"
repeat_region 4348. .4375
/rpt_family="(TA)n"
repeat_region complement(4456. .4494)
/rpt_family="L2"
repeat_region complement(4563. .4977)
/rpt_family="Tigger2"
repeat_region 5199. .5488
/rpt_family="AluSq"
repeat_region complement(5658. .5877)
/rpt_family="THE1C"
repeat_region complement(5877. .6081)
/rpt_family="THE1C"
repeat_region complement(6413. .6734)
/rpt_family="AluJo"
repeat_region 6953. .7269
/rpt_family="AluSx"
repeat_region 7286. .7318
/rpt_family="(CAGG)n"
repeat_region 7498. .7591
/rpt_family="MER45A"
repeat_region complement(7592. .8322)
/rpt_family="L1PA10"
repeat_region 8323. .8405
/rpt_family="MER45A"
repeat_region complement(8434. .8488)
/rpt_family="MER94"
repeat_region 8592. .8886
/rpt_family="AluY"
repeat_region 10059. .10237
/rpt_family="L1ME"
repeat_region 11324. .11367
/rpt_family="AT_rich"
repeat_region complement(11666. .11987)
/rpt_family="LTR16A"
repeat_region complement(12098. .12265)
/rpt_family="HERV16"
repeat_region complement(12345. .12443)
/rpt_family="HERV16"
repeat_region 12530. .12737
/rpt_family="MLT1E"
repeat_region complement(12811. .13097)
/rpt_family="AluSx"
repeat_region complement(13748. .13907)
/rpt_family="HERV16"
repeat_region complement(13908. .14210)
/rpt_family="AluSx"
repeat_region complement(14211. .15723)
/rpt_family="HERV16"
repeat_region complement(15768. .16359)
/rpt_family="HERV16"
repeat_region complement(16612. .16841)
/rpt_family="HERV16"

```
     repeat_region   complement(16903. .17123)
                     /rpt_family="HERV16"
     repeat_region   17127. .17249
                     /rpt_family="FLAM_C"
     repeat_region   complement(17497. .17713)
                     /rpt_family="LTR16A"
     repeat_region   complement(21862. .22326)
                     /rpt_family="L1PA16"
     repeat_region   24395. .24684
                     /rpt_family="AluSx"
     repeat_region   24685. .24707
                     /rpt_family="AT_rich"
     repeat_region   25040. .25098
                     /rpt_family="AT_rich"
     repeat_region   complement(27450. .28648)
                     /rpt_family="MER69B"
     repeat_region   28656. .28788
                     /rpt_family="MIR"
     repeat_region   28789. .29098
                     /rpt_family="AluSx"
     repeat_region   29099. .29162

  Query Match             57.8%;  Score 21.4;  DB 9;  Length 170257;
  Best Local Similarity   80.6%;  Pred. No. 1.6e+02;
  Matches   25;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      6 ggatcccatttttccttagcataacggaagtc 36
          || |||| |||| |||| |||||| |||| |
Db 128767 GGGTCCCTTTTTTCTTATCATAACAGAAGGC 128797


RESULT  13
AL357061
LOCUS       AL357061               70148 bp    DNA     linear   PRI 24-OCT-2001
DEFINITION  Human DNA sequence from clone RP11-694F24 on chromosome 13,
            complete sequence.
ACCESSION   AL357061
VERSION     AL357061.19  GI:16444694
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 70148)
  AUTHORS   Tromans,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 25, 2001 this sequence version replaced gi:16214633.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-694F24 is from the library RPCI-11.3 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-694F24 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-694F24 is at 70148 in this
            sequence. The true right end of clone RP11-789G22 is at 2000 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .70148
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-694F24"
                     /clone_lib="RPCI-11.3"
BASE COUNT    19535 a  15215 c  15251 g  20147 t
ORIGIN


  Query Match             57.3%;  Score 21.2;  DB 9;  Length 70148;
  Best Local Similarity   76.5%;  Pred. No. 1.8e+02;
  Matches   26;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      4 ccggatcccatttttccttagcataacggaagtcc 37
          || || |||||||||| ||||||| || |  |||
Db  41101 CCAGACCCCATTTTCATTAGCATGACTGTTTTCC 41134


RESULT  14
AL353701/c
LOCUS       AL353701              107967 bp    DNA     linear   PRI 30-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-203M2 on chromosome 9, complete
            sequence.
ACCESSION   AL353701
VERSION     AL353701.15  GI:11544478
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 107967)
  AUTHORS   Garner,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 4, 2000 this sequence version replaced gi:11342787.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-203M2 is from the library RPCI-11.1 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
```

RP11-203M2 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-203M2 is at 1 in this sequence. The true left end of clone RP11-738I14 is at 107868 in this sequence.
The true right end of clone RP11-479K20 is at 84404 in this sequence.

FEATURES	Location/Qualifiers
source	1. .107967
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="9"
	/clone="RP11-203M2"
	/clone_lib="RPCI-11.1"
repeat_region	178. .724
	/note="L2 repeat: matches 1850. .2403 of consensus"
repeat_region	864. .904
	/note="5S repeat: matches 2. .42 of consensus"
repeat_region	1407. .1731
	/note="AluY repeat: matches 1. .306 of consensus"
repeat_region	2351. .2487
	/note="AluSx repeat: matches 1. .143 of consensus"
repeat_region	2488. .2801
	/note="AluY repeat: matches 1. .311 of consensus"
repeat_region	2802. .2989
	/note="AluSx repeat: matches 143. .311 of consensus"
repeat_region	3322. .3780
	/note="L1MB6 repeat: matches 5706. .6173 of consensus"
repeat_region	3781. .3888
	/note="HAL1 repeat: matches 53. .166 of consensus"
repeat_region	3974. .4146
	/note="HAL1 repeat: matches 161. .355 of consensus"
repeat_region	4152. .4296
	/note="HAL1 repeat: matches 798. .954 of consensus"
repeat_region	4412. .4732
	/note="HAL1 repeat: matches 1177. .1540 of consensus"
repeat_region	4733. .5042
	/note="AluJb repeat: matches 1. .308 of consensus"
repeat_region	5043. .5097
	/note="HAL1 repeat: matches 1540. .1588 of consensus"
repeat_region	5568. .5716
	/note="L2 repeat: matches 2335. .2516 of consensus"
repeat_region	5772. .5878
	/note="L2 repeat: matches 2641. .2749 of consensus"
repeat_region	6188. .6497
	/note="AluY repeat: matches 1. .311 of consensus"
repeat_region	6732. .6757
	/note="L2 repeat: matches 2634. .2656 of consensus"
repeat_region	6758. .7065
	/note="AluSq repeat: matches 1. .302 of consensus"
repeat_region	7066. .7124
	/note="L2 repeat: matches 2656. .2728 of consensus"
repeat_region	7179. .7351
	/note="AluSg/x repeat: matches 127. .299 of consensus"
repeat_region	13752. .14046
	/note="AluJb repeat: matches 1. .302 of consensus"
repeat_region	14210. .14520
	/note="AluSx repeat: matches 1. .312 of consensus"
repeat_region	14522. .14571
	/note="25 copies 2 mer ac 92% conserved"
repeat_region	14872. .15071
	/note="MER3 repeat: matches 1. .201 of consensus"
repeat_region	15292. .15588
	/note="AluSq repeat: matches 16. .310 of consensus"
repeat_region	16222. .16529
	/note="AluY repeat: matches 1. .307 of consensus"
repeat_region	16594. .16755
	/note="FRAM repeat: matches 0. .161 of consensus"
repeat_region	17073. .17364
	/note="AluSx repeat: matches 1. .293 of consensus"
repeat_region	17721. .17894
	/note="L1MC5 repeat: matches 7505. .7669 of consensus"
repeat_region	17895. .18185
	/note="AluSx repeat: matches 1. .288 of consensus"
repeat_region	18186. .18399
	/note="L1MC5 repeat: matches 7669. .7913 of consensus"
repeat_region	18879. .19203
	/note="L1MC4 repeat: matches 7091. .7447 of consensus"
repeat_region	19204. .19503
	/note="AluSg repeat: matches 1. .299 of consensus"
repeat_region	19504. .19624
	/note="L1MC4 repeat: matches 7447. .7570 of consensus"
repeat_region	19701. .19843
	/note="L1P5 repeat: matches 5316. .5458 of consensus"
repeat_region	19871. .20175
	/note="AluJo repeat: matches 1. .301 of consensus"
repeat_region	20315. .20581
	/note="AluJb repeat: matches 1. .298 of consensus"
repeat_region	20732. .21027
	/note="AluSq repeat: matches 1. .294 of consensus"
repeat_region	21326. .21373
	/note="L1PA6 repeat: matches 6094. .6141 of consensus"
repeat_region	22178. .22297
	/note="AluSq/x repeat: matches 1. .120 of consensus"
repeat_region	22298. .22356
	/note="Alu repeat: matches 256. .308 of consensus"
repeat_region	24152. .24454
	/note="AluY repeat: matches 2. .303 of consensus"
repeat_region	25128. .25312
	/note="MER8 repeat: matches 21. .214 of consensus"
repeat_region	25330. .25517
	/note="AluJb repeat: matches 108. .294 of consensus"
repeat_region	25706. .26001
	/note="AluSx repeat: matches 1. .301 of consensus"
repeat_region	26146. .26357
	/note="MER20 repeat: matches 7. .218 of consensus"
repeat_region	26399. .26685
	/note="AluSx repeat: matches 2. .288 of consensus"
repeat_region	26732. .26840
	/note="L1MC4 repeat: matches 7859. .7976 of consensus"
repeat_region	26848. .27150
	/note="AluJb repeat: matches 1. .301 of consensus"
repeat_region	27783. .28083
	/note="AluJo repeat: matches 1. .298 of consensus"
repeat_region	28086. .28394
	/note="AluSp repeat: matches 1. .312 of consensus"
repeat_region	28431. .28722
	/note="AluSg repeat: matches 1. .295 of consensus"
repeat_region	29190. .29504
	/note="AluSx repeat: matches 1. .311 of consensus"
repeat_region	29685. .29734
	/note="L2 repeat: matches 2648. .2697 of consensus"
repeat_region	29759. .30061
	/note="AluJb repeat: matches 1. .301 of consensus"
repeat_region	30161. .30232
	/note="AluJ/FRAM repeat: matches 231. .302 of consensus"
repeat_region	31184. .31483
	/note="AluSx repeat: matches 1. .303 of consensus"
repeat_region	33583. .33767
	/note="AluSc repeat: matches 1. .184 of consensus"
repeat_region	33812. .34357
	/note="L2 repeat: matches 1706. .2312 of consensus"
misc_feature	35355. .36871
	/note="CpG island"
	/evidence=not_experimental
repeat_region	35383. .35638
	/note="AluJo repeat: matches 55. .312 of consensus"
repeat_region	36676. .36817
	/note="AluJo repeat: matches 6. .150 of consensus"
repeat_region	36818. .37125
	/note="AluSq repeat: matches 1. .309 of consensus"
repeat_region	37126. .37287
	/note="AluJo repeat: matches 150. .303 of consensus"
repeat_region	37322. .37614
	/note="AluSx repeat: matches 1. .294 of consensus"
repeat_region	38031. .38059
	/note="ORSL repeat: matches 277. .309 of consensus"

```
     repeat_region   38060. .38354
                     /note="AluY repeat: matches 1. .296 of consensus"
     repeat_region   38355. .38438
                     /note="ORSL repeat: matches 214. .277 of consensus"
     repeat_region   38785. .38999
                     /note="AluSx repeat: matches 1. .215 of consensus"
     repeat_region   39026. .39083
                     /note="29 copies 2 mer aa 77% conserved"
     repeat_region   39090. .39291
                     /note="L1M4 repeat: matches 4536. .4727 of consensus"
     repeat_region   39294. .39379
                     /note="MER6A repeat: matches 2. .90 of consensus"

  Query Match             57.3%;  Score 21.2;  DB 9;  Length 107967;
  Best Local Similarity   76.5%;  Pred. No. 1.9e+02;
  Matches   26;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 gcgccggatcccatttttccttagcataacggaag 34
           | ||| |  | ||||||||||  ||||||||||
Db   85590 GAGCCAGCACACATTTTCCTTCCTATAACGGAAG 85557


RESULT  15
AP003009/c
LOCUS       AP003009              339681 bp    DNA     linear   BCT 15-MAY-2001
DEFINITION  Mesorhizobium loti DNA, complete genome, section 16/21.
ACCESSION   AP003009 BA000012
VERSION     AP003009.2  GI:14026063
KEYWORDS    .
SOURCE      Mesorhizobium loti (strain:MAFF303099) DNA.
  ORGANISM  Mesorhizobium loti
            Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Phyllobacteriaceae; Mesorhizobium.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
  JOURNAL   DNA Res. 7 (6), 331-338 (2000)
  MEDLINE   21082930
REFERENCE   2  (bases 1 to 339681)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     On May 11, 2001 this sequence version replaced gi:11994984.
FEATURES             Location/Qualifiers
     source          1. .339681
                     /organism="Mesorhizobium loti"
                     /strain="MAFF303099"
                     /db_xref="taxon:381"
     gene            complement(82. .420)
                     /gene="mll6350"
     CDS             complement(82. .420)
                     /gene="mll6350"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52662.1"
                     /db_xref="GI:14026064"
                     /translation="MPLLAFDLLARIALPKVGLPLERVNLEACSLTAWPHDGLRAEGQ
                     PVHVKSWQCRLWRSLLVHRHPVLNEMRSIENMVRTILREACINWARGREPPSPVVCRS
                     WPALTPRCWK"
     gene            447. .593
                     /gene="msr6351"
     CDS             447. .593
                     /gene="msr6351"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52663.1"
                     /db_xref="GI:14026065"
                     /translation="MHALVHAADIQDCDGGVMVMAKLFGLDPFLFKLYAGGGHSRGRS
                     SKRG"
     gene            complement(733. .1464)
                     /gene="mll6352"
     CDS             complement(733. .1464)
                     /gene="mll6352"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="BAB52664.1"
                     /db_xref="GI:14026066"
                     /translation="MRTWAMSEVARDPLYRCHRFPAEIIAHAVWLYFRFPLSLRMVED
                     MLAARGIIVTHQTIRSWAEKFGRHFAREIKRRSAGCLGDKWHLDECVVAINGKKQWLW
                     RAVDQDGLVLEVLVQSRRNAKAAKRLMRKLLKAQGRTPRVMITDKLRSYDAARRDLMP
                     GVEHRSHKGLNNRAENSHQPTRRRERTMKRFKSARQLQRFVSIHDPIANLFHFPRNTL
                     SSAQHRDLRNAAMQIWNKIACLAAA"
     gene            complement(1080. .1748)
                     /gene="mll6353"
     CDS             complement(1080. .1748)
                     /gene="mll6353"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="BAB52665.1"
                     /db_xref="GI:14026067"
                     /translation="MVEACATYKNHRFPIEIVAHAVWLYYRFGLSLRDVEEMLLERGI
                     VVSYETIRRWGKKHSPDYARRLRRKAPSKDDVWHLDEVAGHRKVNRRLMKMCEHGPCQ
                     KLLVIRFIVATAFPPRSLRTRYGSTFAFRSACAWLRTCWRPAASSSRIRPFEAGRRNS
                     GGISPGRSSDGQPAAWATNGISTNVWWPSMARSSGSGVPSIRTAWSSKCWCKAAEMPR
                     RQSV"
     gene            complement(1766. .2200)
                     /gene="mll6354"
     CDS             complement(1766. .2200)
                     /gene="mll6354"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52666.1"
                     /db_xref="GI:14026068"
                     /translation="MVARHATIAAEAVERTRTPPSRVRRQAEHAPVSAKELAQALPED
                     AWHMVTWREGSKASLASRFATVRVRPAHRDYWRSIPHAEEWFLIEWPDGEVEPTKYWL
                     STLPGKTRLADLVDQAKMRWRIERDYQELKQGCCHIDFGLTV"
     gene            complement(1939. .2931)
                     /gene="mll6355"
     CDS             complement(1939. .2931)
                     /gene="mll6355"
                     /codon_start=1
                     /transl_table=11
                     /product="transposase"
                     /protein_id="BAB52667.1"
                     /db_xref="GI:14026069"
                     /translation="MLHMEGAMDLQDGIETSESRFAAYVETLASVLGHADRVAPLKAY
                     CTGLLLPRERKSVEPMAARVEPGRVQAAHQSLHHFVAKADWSDDAVLSRVREQVLPAL
                     ERQGQIRAWIVDDTGFPKKGKHSVGVARQYCGQLGKQDNCQVAVSLSVATEQASLPIA
                     YQLYLPESWANDPDRRAKTGVPENVVFRTKPEIALAQIRAALEAGVSRGVVLADAGYG
                     NDTAFRTGLTKVGLTYVVGVQSSIRLWSPGTQPLPPKPWSGRGPRPRAFGGRLNMRRS
                     QPRNWRKRCPRTPGTWLLGARAPRRRLPHALPPCGFGRPIATTGAPSPTLRNGS"
     gene            complement(3398. .3616)
                     /gene="msl6356"
     CDS             complement(3398. .3616)
                     /gene="msl6356"
                     /note="unknown protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB52668.1"
                     /db_xref="GI:14026070"
                     /translation="MLSGMPPCRLLRTMMMAGATIDLLAGQAFLVAGFQIETRTARHH
```

RSAAKRDPRSPGLRNLTNSARPNADRSL"

gene 4196. .6634
/gene="mlr6358"

CDS 4196. .6634
/gene="mlr6358"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB52669.1"
/db_xref="GI:14026071"
/translation="MADTSAPIHRQRVLVYIGDEKMTGIGRSGKSRIGDAGAGSSSSR
SQSGLASGEGSQGFGSVLERMRSEGGKGSPSAGAARGAHDDLGGKSSGPSVPPHRIPA
AAPRAAAGPRAAGSRRRSVSSTSDGEALAPVASQPSTSGVTVPSSSSGEVINETKRRM
CSLIDEIQVLEEKSLHSGKPMEGFRDELDQLANAGAELLECYASLPKDEARSTLSDDD
VRHYRADVIHRADSANELAHTIFKKEEQNRKRMEDILHNLVKSNATLDQLSRSAPSMA
KYPVACVERWDKKVQRHELMRAVCVATADLPSSTAEMRREEEAALRLHSGWVVNSKIM
QLQSRINLAQVNMAPQAIMLEAPIRELLIGDNGQTRLLGNFIGEVFPAFVSTAEAVLR
KEGRPLDAEHCAVLEGVVERLADFASASHSVVDRLRDHQAGAGLPLELLTRIVDDAWI
ASDEVTDLLALQPKPPAPDLTADAAAAMPTGTAGKTMVAESTTAPRKKGKDKSKQAAG
AGSSAAGRPEPQVAAAVSDPAPAAKVLVRTDLGTKKLVSAEEAQNSASSAAARLAILQ
NPTTGENTKSLVARSTELLKFDLPAQQSLISRARSGKPEDVEHAVGLVVQRLQTQAVE
MEPCLPALEASRLSKEGKMDDVHRLTVQFRDMLANVRGLAKAWDEQKPDMIMDCMKRY
PFPSQNHLEQLRSARELTPADPPCALNPEPSNLFEIRLQPKALSNRTLPSPMWVHIHT
TRKIYAQQLATLDSAEFAACHVKSNEQRGHNRQWQTSRAKKGYDNAIHRGKLTPAFCK
SLLSDRADMRNGPA"

gene complement(6985. .7716)
/gene="mll6359"

CDS complement(6985. .7716)
/gene="mll6359"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB52670.1"
/db_xref="GI:14026072"
/translation="MRTWAMSEVARDPLYRCHRFPAEIIAHAVWLYFRFPLSLRMVED
MLAARGIIVTHQTIRSWAEKFGRHFAREIKRRSAGCLGDKWHLDECVVAINGKKQWLW
RAVDQDGLVLEVLVQSRRNAKAAKCLMRKLLKAQGRTPRVMITDKLRSYDAARRDLMP
GVEHRSHKGLNNRAENSHQPTRRRERTMKRFKSARQLQRFVSIHDPIANLFHFPRNTL
SSAQHRDLRNAAMQIWNKIACLAAA"

gene 8086. .17256
/gene="mlr6361"

CDS 8086. .17256
/gene="mlr6361"
/note="similar to shikimate kinase"
/codon_start=1
/transl_table=11
/protein_id="BAB52671.1"
/db_xref="GI:14026073"
/translation="MTGVNRSGRSRTEKRVGRTGEASSSRSQSSLAPGAGSQSFDTVV
EQMRSVEINRRTPAVPMERGADDALHDRPSGPSVGIHGTLDAARHAAAAPRAAGPRRR
SVPSMLEGETLAPVASQPPTSGVAAPICLSGDEIKEKRRMDSLLTLVGMRFNAARQDP
GDEKLLDQVINAGSAALHHYKSLSPALARTIVDDNQLITCRYLVWDAVQEYSRLAEPT
IHKLQKKKAEEEAIGLTQSTATPESVVEAHAACVKWWEKKVGHCERGRAAWRATASLP
SATAKMRQEEEATLRLATGSVMHSKLVYLQSRIDLARAKIEPQASTFDAPVREILMVE
NGRVRLLEAFSGDVLPAFLSTQNAVLSSGGHPLDADHCAVLEGVIERLSEFASELQVI
RTNLNDHQTSADLPLELLSQIVEGAWIVGHEVMHLLELQPKPPAVALPPDAGAAIPAE
TAGKSRVAKGTLALQPKAPALIASPADAGATIQIDTVGKSPVAKGTLALQPNWPANIA
PPADMGAGIPADTAGKAAVAGAAARRKRKGKGKPAVGAGSSAPGRASQVAAVDSDTAP
ADKVLVLSDMGTKLVSAEEARASASAGAQSAKQAPPSMEALTQERQVKPEDARYVAET
VVERLQTQPAEMRACPAALDEPRRPGLLAPAQVPEVQDKTVRLKGMLNQAQELAKALN
SQSSHAPGQSSQSSDAGVERTHSGPQDRMSYARVVRGADEALGDKSADYSIAPRGALA
GSRGATSGPRHGSPQGRHGEAMGRQESSLRLERSWLGDQEPSRSQPDTLARHDQSQSR
PVAGPERRRFDHVAAQSFDDALLNDALHKIFFAHDINEFSYAVVENDKRLQSQAHRAT
FLRKAASEFTRHFVPRWEDDVRAGNTWGFATSCNATSREPGGQAGMEACRAMAAQVSR
LGRALNDVESKTLSLFVLSFSRHPRLAECRNGMITIAKLFHEQPGALSELNGQSLALL
VNGFSKWPEQENARKATVAVAAEIRRRPAGLSDFDPQNLGNVVNGLSKWPLEEECCGV
IDAIAGKVCGRAGEKAGLSGFNSQAFANLVNGFSKAREPGNSHQATIVIADAVCRLAD
EIVIADSIGRLADEKAGRTGWYPQALANLVNGFSKWPLDCRGAIVAIARQVAGGAGKL
PVFDAQALANLVNGFSKWPEDCRGATIAIAAEVCRLADLPKAGLSGFDPQHLANLVNG
FSKLQAAECSRATVAIAGEVCHRAGRRKGLSDFDVQHLANLVKGFSKWPQEVNLGYAT
NVVAGEVLRCAEPLSEFLPLHLASLVNGFSKWPRWENSHQATVAIAGEVRHRADQLSG
FEPQELANLVNGFSKWPEDCRGAIVAIARAVPRRAGRLSGFNGQDLANLVNGFSKWPL
DCREAMVAIARAVPRRAGRLSGFNGQDLANLVNGFSKWPQEENSRQATVAIAGEVLRS
TNPPPDFSSQYLANLVNGFSKWPEEEALRQAAAAIAGEVIRSANRLPDFTSQHLANLV
NGFSKWPQEEKSRQATAAIAGEVLRGAERLSGVNPQQLANLLNGFDRFAEEEACSQAI
LEIARRLGQAGQPFRHFAAPGLSSIANSLARGILRSEDAGEIAEAALLKDRLHKLAHY
LHYASDRLEEADAVGVTSILKALAKAQLYDDLGSLAGAGLNRLAELHRAPGFAHENNL
ETMGNLCAALLPLARSPRKQLLWHRRPALNLLNDIQPIVEHKIEAHLRASDAERARGP

Query Match 57.3%; Score 21.2; DB 1; Length 339681;
Best Local Similarity 76.5%; Pred. No. 2e+02;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy      1 gcgccggatcccatttttccttagcataacggaaag 34
          ||| |||||||||||   |||| ||||||  || |
Db 232500 GCGGCGGATCCCATCGGCCTTTGCATAATCGACG 232467
```

Search completed: July 31, 2002, 19:32:00
Job time: 8466 sec

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 17:09:18 ; Search time 2686.26 Seconds
(without alignments)
221.075 Million cell updates/sec

Title: US-09-824-567-3
Perfect score: 44
Sequence: 1 ataagaatgcggccgccacc.....gcaagatatcagtgggaatc 44

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	24.6	55.9	493	10	BE442509	BE442509 925021G12
2	24.6	55.9	505	10	BI723400	BI723400 1031066G1
3	24.6	55.9	511	10	BI720753	BI720753 1031051H0
4	24.6	55.9	521	10	BI720615	BI720615 1031051A0
5	24.6	55.9	613	10	BM000737	BM000737 1031090H0
6	24.6	55.9	624	10	BM001235	BM001235 1031094A1
7	24.6	55.9	626	10	BI723401	BI723401 1031066G1
8	24.6	55.9	638	10	BI727233	BI727233 1031090H0
c 9	24.4	55.5	437	9	AI758849	AI758849 ty16b10.x
c 10	24	54.5	291	9	BB229997	BB229997 BB229997
11	23.6	53.6	648	12	AG059299	AG059299 Pan trogl
12	23.4	53.2	648	12	AG075580	AG075580 Pan trogl
13	23.4	53.2	676	12	AG057076	AG057076 Pan trogl
14	23.4	53.2	694	12	AG101006	AG101006 Pan trogl
15	23.2	52.7	452	9	AW547423	AW547423 L0022A10-
16	23.2	52.7	680	12	AG078608	AG078608 Pan trogl
17	23	52.3	239	9	AV370416	AV370416 AV370416
c 18	23	52.3	554	10	BE500188	BE500188 WHE0980_F
c 19	23	52.3	572	10	BM135137	BM135137 WHE0454_D
c 20	23	52.3	582	10	BE417948	BE417948 SCL013.C0
21	23	52.3	644	12	AG106142	AG106142 Pan trogl
22	23	52.3	654	12	AG039910	AG039910 Pan trogl
23	23	52.3	668	12	AG122582	AG122582 Pan trogl
24	23	52.3	690	12	AG071800	AG071800 Pan trogl
c 25	23	52.3	690	12	AG138371	AG138371 Pan trogl
26	23	52.3	695	12	AG125018	AG125018 Pan trogl
27	23	52.3	715	12	AG093362	AG093362 Pan trogl
28	23	52.3	746	10	BI647012	BI647012 603278786
29	22.8	51.8	301	10	BG655950	BG655950 ib35b03.y
30	22.8	51.8	413	9	AI283230	AI283230 qk50d06.x
31	22.8	51.8	486	9	AA948385	AA948385 on52a11.s
32	22.8	51.8	648	12	AG033798	AG033798 Pan trogl
33	22.8	51.8	719	12	AG051700	AG051700 Pan trogl
34	22.8	51.8	727	10	BI602851	BI602851 603250319
35	22.8	51.8	759	10	BI192836	BI192836 602945083
36	22.8	51.8	804	10	BI544724	BI544724 603242683
37	22.6	51.4	720	12	AG052860	AG052860 Pan trogl
38	22.4	50.9	419	9	AV646405	AV646405 AV646405
c 39	22.4	50.9	429	9	BB820621	BB820621 BB820621
40	22.4	50.9	524	12	AQ779060	AQ779060 HS_3084_B
41	22.4	50.9	576	12	AQ805313	AQ805313 HS_3214_A
c 42	22.4	50.9	619	10	BI148691	BI148691 602912125
c 43	22.4	50.9	632	11	BC019210	BC019210 Mus muscu
44	22.4	50.9	647	12	AG036590	AG036590 Pan trogl
45	22.4	50.9	650	12	AG036056	AG036056 Pan trogl

ALIGNMENTS

RESULT 1
BE442509
LOCUS BE442509 493 bp mRNA linear EST 25-JUL-2000
DEFINITION 925021G12.x1 C. reinhardtii CC-2290, normalized, Lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BE442509
VERSION BE442509.1 GI:9442025
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 493)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Elizabeth H. Harris
DCMB Box 91000
Duke University
Durham, NC 27708-1000, USA
Tel: 919 613 8164
Fax: 919 613 8177
Email: chlamy@duke.edu.
FEATURES Location/Qualifiers
source 1. .493
/organism="Chlamydomonas reinhardtii"
/strain="CC-2290 wild type mt- S1 D2"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-2290, normalized, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library was constructed by John Davies and Jeffrey McDermott. RNA was isolated from strain CC-2290 (Minnesota isolate of C. reinhardtii) grown to mid-log phase in TAP (acetate containing) medium in the light. PolyA mRNA was purified, and cDNA was synthesized and directionally cloned into lambda ZAP II (Stratagene) in

the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 129 a 145 c 111 g 108 t
ORIGIN

Query Match 55.9%; Score 24.6; DB 10; Length 493;
Best Local Similarity 76.9%; Pred. No. 47;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaagatatcagtgg 39
Db 384 ACATGAATGCGGCCTCAACTACGCGCAAGTTATTGGTGG 422

RESULT 2
BI723400
LOCUS BI723400 505 bp mRNA linear EST 19-SEP-2001
DEFINITION 1031066G11.x2 C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI723400
VERSION BI723400.1 GI:15699079
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 505)
AUTHORS Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1031
JOURNAL Unpublished (2001)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES Location/Qualifiers
source 1. .505
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoRI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."
BASE COUNT 129 a 137 c 119 g 120 t
ORIGIN

Query Match 55.9%; Score 24.6; DB 10; Length 505;
Best Local Similarity 76.9%; Pred. No. 48;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaagatatcagtgg 39
Db 445 ATATGAATGCGGCCTCAGCTACGCGCAAGTTATTGGTGG 483

RESULT 3
BI720753
LOCUS BI720753 511 bp mRNA linear EST 19-SEP-2001
DEFINITION 1031051H01.x1 C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI720753
VERSION BI720753.1 GI:15696448
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 511)
AUTHORS Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1031
JOURNAL Unpublished (2001)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES Location/Qualifiers
source 1. .511
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoRI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."
BASE COUNT 129 a 143 c 121 g 118 t
ORIGIN

Query Match 55.9%; Score 24.6; DB 10; Length 511;
Best Local Similarity 76.9%; Pred. No. 48;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaagatatcagtgg 39
Db 441 ATATGAATGCGGCCTCAGCTACGCGCAAGTTATTGGTGG 479

RESULT 4
BI720615
LOCUS BI720615 521 bp mRNA linear EST 19-SEP-2001
DEFINITION 1031051A01.x1 C. reinhardtii CC-1690, Stress II (normalized),

```
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI720615
VERSION     BI720615.1  GI:15696310
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                     Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
                     Research 6: 791-806."
BASE COUNT      130 a    147 c    123 g    121 t
ORIGIN


  Query Match           55.9%;  Score 24.6;  DB 10;  Length 521;
  Best Local Similarity  76.9%;  Pred. No. 48;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtgg 39
          ||| |||||||||| |  | | ||||||| |||  ||||
Db    443 ATATGAATGCGGCCTCAGCTACGCGCAAGTTATTGGTGG 481


RESULT   5
BM000737
LOCUS       BM000737                613 bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  1031090H07.x2 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BM000737
VERSION     BM000737.1  GI:16435517
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                     Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
                     Research 6: 791-806."
BASE COUNT      158 a    182 c    138 g    135 t
ORIGIN


  Query Match           55.9%;  Score 24.6;  DB 10;  Length 613;
  Best Local Similarity  76.9%;  Pred. No. 51;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtgg 39
          ||| |||||||||| |  | | ||||||| |||  ||||
Db    445 ATATGAATGCGGCCTCAGCTACGCGCAAGTTATTGGTGG 483


RESULT   6
BM001235
LOCUS       BM001235                624 bp    mRNA    linear   EST 25-OCT-2001
DEFINITION  1031094A10.x2 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BM001235
VERSION     BM001235.1  GI:16436015
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
```

```
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                     Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
                     Research 6: 791-806."
BASE COUNT      162 a    187 c    139 g    136 t
ORIGIN


  Query Match             55.9%;  Score 24.6;  DB 10;  Length 624;
  Best Local Similarity   76.9%;  Pred. No. 51;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgcgcaagatatcagtgg 39
           ||| |||||||||| |  | | ||||||| |||  ||||
Db     445 ATATGAATGCGGCCTCAGCTACGCGCAAGTTATTGGTGG 483


RESULT   7
BI723401
LOCUS       BI723401                  626 bp    mRNA    linear   EST 19-SEP-2001
DEFINITION  1031066G11.x3 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI723401
VERSION     BI723401.1  GI:15699080
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .626
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                     Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
                     Research 6: 791-806."
BASE COUNT      162 a    187 c    140 g    137 t
ORIGIN


  Query Match             55.9%;  Score 24.6;  DB 10;  Length 626;
  Best Local Similarity   76.9%;  Pred. No. 51;
  Matches   30;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgcgcaagatatcagtgg 39
           ||| |||||||||| |  | | ||||||| |||  ||||
Db     446 ATATGAATGCGGCCTCAGCTACGCGCAAGTTATTGGTGG 484


RESULT   8
BI727233
LOCUS       BI727233                  638 bp    mRNA    linear   EST 19-SEP-2001
DEFINITION  1031090H07.x1 C. reinhardtii CC-1690, Stress II (normalized),
            Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI727233
VERSION     BI727233.1  GI:15702928
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 638)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1031
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .638
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
                     ), Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; Stress condition II library, constructed by John
                     Davies and Jeffrey McDermott, combines cDNAs from CC-1690
                     cells grown to mid-log phase in TAP (NH4+ - containing)
                     and shifted to TAP - NO3- (24hrs); H2 production
                     conditions (0, 12hr, 24hr) see Melis et al.,(2000) Plant
                     Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
                     sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     Zap II (Stratagene) in the EcoRI (5') and XhoRI (3')
                     sites. pBluescript II SK- plasmids were excised from the
                     lambda ZAP clones by superinfection with ExAssist
                     (Stratagene) phage. The library was normalized using
                     method 4 described in Bonaldo et al., (1996) Genome
```

Research 6: 791-806."
BASE COUNT 165 a 191 c 142 g 140 t
ORIGIN

Query Match 55.9%; Score 24.6; DB 10; Length 638;
Best Local Similarity 76.9%; Pred. No. 52;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaagatatcagtgg 39
||| |||||||||| | | | ||||||| ||| ||||
Db 445 ATATGAATGCGGCCTCAGCTACGCGCAAGTTATTGGTGG 483

RESULT 9
AI758849/c
LOCUS AI758849 437 bp mRNA linear EST 16-DEC-1999
DEFINITION ty16b10.x1 NCI_CGAP_Ut3 Homo sapiens cDNA clone IMAGE:2279227 3', mRNA sequence.
ACCESSION AI758849
VERSION AI758849.1 GI:5152574
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 437)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1863 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 401.
FEATURES Location/Qualifiers
source 1. .437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2279227"
/clone_lib="NCI_CGAP_Ut3"
/tissue_type="poorly-differentiated endometrial adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.45 kb. Life Technologies catalog #: 11541-018"
BASE COUNT 149 a 77 c 77 g 130 t 4 others
ORIGIN

Query Match 55.5%; Score 24.4; DB 9; Length 437;
Best Local Similarity 72.1%; Pred. No. 54;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2 taagaatgcggccgccaccatgcgcaagatatcagtgggaatc 44
|| ||| |||| ||||| ||||||| | ||||| ||||
Db 158 TANAAATTGGGCCCCCACCCAACGCAAGAACTGAGTGGAAATC 116

RESULT 10
BB229997/c
LOCUS BB229997 291 bp mRNA linear EST 03-JUL-2000
DEFINITION BB229997 RIKEN full-length enriched, 3 days neonate thymus Mus musculus cDNA clone A630024I16 3', mRNA sequence.
ACCESSION BB229997
VERSION BB229997.1 GI:8900642
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 291)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .291
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A630024I16"
/clone_lib="RIKEN full-length enriched, 3 days neonate thymus"
/tissue_type="thymus"
/dev_stage="3 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. lst strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a

modified pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 59 a 73 c 76 g 83 t
ORIGIN

Query Match 54.5%; Score 24; DB 9; Length 291;
Best Local Similarity 75.0%; Pred. No. 67;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaagatatcagtggg 40
Db 137 ATAAAAGTGTGGCCGCCTCCTTGGGCAAGCAATCATGGGG 98

RESULT 11
AG059299
LOCUS AG059299 648 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-046H05.R, genomic survey sequence.
ACCESSION AG059299
VERSION AG059299.1 GI:16610529
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-046H05.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 648)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1. .648
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-046H05.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 162 a 208 c 150 g 128 t
ORIGIN

Query Match 53.6%; Score 23.6; DB 12; Length 648;
Best Local Similarity 76.3%; Pred. No. 1.2e+02;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 4 agaatgcggccgccaccatgcgcaagatatcagtggga 41
Db 20 ATAAGGCCGCCGCTAGCATGCTCGAGAGCTCAGTGGGA 57

RESULT 12
AG075580
LOCUS AG075580 648 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-068I10.R, genomic survey sequence.
ACCESSION AG075580
VERSION AG075580.1 GI:16627382
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-068I10.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 648)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1. .648
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-068I10.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 184 a 143 c 145 g 176 t
ORIGIN

Query Match 53.2%; Score 23.4; DB 12; Length 648;
Best Local Similarity 73.2%; Pred. No. 1.5e+02;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 4 agaatgcggccgccaccatgcgcaagatatcagtgggaatc 44
Db 29 ATAGGGCGGCCGCTAGCATGCTCGAGAGCTCAATGGGAACC 69

RESULT 13
AG057076
LOCUS AG057076 676 bp DNA linear GSS 02-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-043I21.R, genomic survey sequence.
ACCESSION AG057076
VERSION AG057076.1 GI:16594535
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-043I21.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 676)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,

```
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-043I21.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      168 a    161 c    134 g    213 t
ORIGIN


  Query Match             53.2%;  Score 23.4;  DB 12;  Length 676;
  Best Local Similarity   73.2%;  Pred. No. 1.5e+02;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 agaatgcggccgccaccatgcgcaagatatcagtgggaatc 44
          | |  |||||||| | ||||| | |||  ||||||  ||||
Db     34 ATAGGGCGGCCGCTAGCATGCTCGAGAGCTCAGTGCTAATC 74


RESULT  14
AG101006
LOCUS       AG101006                694 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-103M03.R, genomic survey sequence.
ACCESSION   AG101006
VERSION     AG101006.1  GI:16721523
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-103M03.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 694)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .694
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-103M03.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      185 a    135 c    132 g    239 t      3 others
ORIGIN


  Query Match             53.2%;  Score 23.4;  DB 12;  Length 694;
  Best Local Similarity   73.2%;  Pred. No. 1.5e+02;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 agaatgcggccgccaccatgcgcaagatatcagtgggaatc 44
          | |  |||||||| | ||||| | |||  ||||||  ||||
Db     34 ATAGGGCGGCCGCTAGCATGCTCGAGAGCTCACTGGGTATC 74


RESULT  15
AW547423
LOCUS       AW547423                452 bp    mRNA    linear   EST 31-AUG-2000
DEFINITION  L0022A10-3 NIA Mouse E12.5 Female Mesonephros and Gonads cDNA
            Library Mus musculus cDNA clone L0022A10 3', mRNA sequence.
ACCESSION   AW547423
VERSION     AW547423.1  GI:7192837
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
            ,M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraja,R., Doi,H., Wood,W.H.
            III, Becker,K.G. and Ko,M.S.H.
  TITLE     Genome-wide expression profiling of mid-gestation placenta and
            embryo using a 15,000 mouse developmental cDNA microarray
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
  MEDLINE   20381348
COMMENT     Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: L0022  row: A  column: 10
            Seq primer: -21M13 Forward
            High quality sequence stop: 452
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .452
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:L0022A10-3"
                     /db_xref="taxon:10090"
                     /clone="L0022A10"
                     /clone_lib="NIA Mouse E12.5 Female Mesonephros and Gonads
                     cDNA Library"
                     /sex="female"
                     /dev_stage="12.5dpc"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1 (Gibco/BRL Life Technology);
                     Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
                     2 Mesonephros. The double-stranded cDNA  was synthesized
                     by Gibco's kit with an Oligo(dT) primer [NotI
                     primer-adapter from GibcoBRL]
                     [5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from
                     3.42ug of  total RNA . The double-stranded cDNAs were
                     treated with T4 DNA polymerase  and purified by
                     ethanol-precipitation. The cDNAs were ligated to
                     Lone-linker LL-Sal3 (include Sal1 sequence). The cDNAs
                     were purified by  phenol/chloroform and separated from
                     free linkers by Centricon 100.  Then, cDNAs were amplified
```

by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang."

BASE COUNT 107 a 154 c 127 g 64 t
ORIGIN

Query Match 52.7%; Score 23.2; DB 9; Length 452;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy      9 gcggccgccaccatgcgcaagatatcagtgggaatc 44
          | ||| || |||||||||   ||||||| |||| ||
Db    411 GGGGCGGCGACCATGCGCTCAATATCAGGGGGACTC 446
```

Search completed: July 31, 2002, 18:59:48
Job time: 6630 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 17:11:44 ; Search time 84.08 Seconds
(without alignments)
128.543 Million cell updates/sec

Title: US-09-824-567-3
Perfect score: 44
Sequence: 1 ataagaatgcggccgccacc.....gcaagatatcagtgggaatc 44

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	22	50.0	414	2	US-08-766-439-20	Sequence 20, Appl
	2	22	50.0	414	2	US-08-766-439-21	Sequence 21, Appl
c	3	22	50.0	1274	2	US-08-766-439-28	Sequence 28, Appl
	4	22	50.0	1274	2	US-08-766-439-29	Sequence 29, Appl
c	5	22	50.0	1327	2	US-08-766-439-26	Sequence 26, Appl
	6	22	50.0	1327	2	US-08-766-439-27	Sequence 27, Appl
c	7	22	50.0	1354	2	US-08-766-439-24	Sequence 24, Appl
	8	22	50.0	1354	2	US-08-766-439-25	Sequence 25, Appl
	9	20.6	46.8	45	3	US-08-974-022-16	Sequence 16, Appl
	10	20.6	46.8	45	4	US-08-795-445A-16	Sequence 16, Appl
	11	20.6	46.8	45	4	US-08-795-447A-16	Sequence 16, Appl
	12	20.6	46.8	45	4	US-08-974-186-16	Sequence 16, Appl
	13	20.6	46.8	45	4	US-08-795-446B-16	Sequence 16, Appl
c	14	20.6	46.8	2412	1	US-08-158-232-9	Sequence 9, Appli
c	15	20.6	46.8	2412	1	US-08-304-626-9	Sequence 9, Appli
c	16	20.6	46.8	2412	1	US-08-316-301A-11	Sequence 11, Appl
c	17	20.6	46.8	2412	1	US-08-611-928-9	Sequence 9, Appli
c	18	20.6	46.8	2412	3	US-09-173-891-9	Sequence 9, Appli
c	19	20.6	46.8	2412	4	US-09-076-137-11	Sequence 11, Appl
c	20	20.6	46.8	2412	5	PCT-US92-03624-11	Sequence 11, Appl
	21	20	45.5	35	5	PCT-US96-10905-35	Sequence 35, Appl
c	22	20	45.5	1491	4	US-09-058-947A-3	Sequence 3, Appli
c	23	20	45.5	1502	4	US-08-868-373-11	Sequence 11, Appl
c	24	20	45.5	1807	4	US-09-058-947A-2	Sequence 2, Appli
c	25	20	45.5	3722	4	US-09-058-947A-1	Sequence 1, Appli
	26	19.6	44.5	39	3	US-08-814-052-55	Sequence 55, Appl
	27	19.6	44.5	60	5	PCT-US94-08052-5	Sequence 5, Appli
c	28	19.6	44.5	2249	3	US-08-814-052-19	Sequence 19, Appl
c	29	19.6	44.5	2300	3	US-08-814-052-18	Sequence 18, Appl
	30	19.4	44.1	75	4	US-09-459-956-20	Sequence 20, Appl
	31	19.4	44.1	15378	3	US-08-785-420-1	Sequence 1, Appli
	32	19.2	43.6	1875	1	US-08-453-956-14	Sequence 14, Appl
	33	19.2	43.6	1875	1	US-08-086-631-14	Sequence 14, Appl
	34	19.2	43.6	1875	2	US-08-452-930-14	Sequence 14, Appl
	35	19.2	43.6	1875	5	PCT-US93-08174-14	Sequence 14, Appl
c	36	19.2	43.6	2214	4	US-08-943-731-57	Sequence 57, Appl
	37	19.2	43.6	2379	4	US-08-797-358B-2	Sequence 2, Appli
c	38	19.2	43.6	18609	4	US-08-943-731-1	Sequence 1, Appli
	39	19	43.2	44	1	US-08-106-078-7	Sequence 7, Appli
	40	19	43.2	44	1	US-08-591-492-7	Sequence 7, Appli
c	41	19	43.2	1501	2	US-08-145-658D-24	Sequence 24, Appl
c	42	18.8	42.7	414	2	US-08-766-439-22	Sequence 22, Appl
	43	18.8	42.7	414	2	US-08-766-439-23	Sequence 23, Appl
c	44	18.8	42.7	560	2	US-08-330-272-3	Sequence 3, Appli
c	45	18.8	42.7	560	5	PCT-US95-13663-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT   1
US-08-766-439-20/c
; Sequence 20, Application US/08766439
; Patent No. 5922538
;  GENERAL INFORMATION:
;    APPLICANT:  HAZEL, JAMES WILLIAM
;    APPLICANT:  JENSEN, MARK ANTON
;    TITLE OF INVENTION:  GENETIC MARKERS AND METHODS FOR
;    TITLE OF INVENTION:  THE DETECTION OF LISTERIA
;    TITLE OF INVENTION:  MONOCYTOGENES AND LISTERIA SPP.
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.50 INCH DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,439
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/745,228
;      FILING DATE:  NOVEMBER 8, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  MD-1065-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  414 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      STRAIN:  L MONO - 647 - PREMARKER
US-08-766-439-20


  Query Match                 50.0%;  Score 22;  DB 2;  Length 414;
```

```
  Best Local Similarity   73.7%;  Pred. No. 2.6;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtg 38
          | ||| |||||||||  |  ||||||||  |||  |||
Db    131 AGAAGCATGCGGCCGAGATAATGCGCAACTTATTTGTG 94


RESULT   2
US-08-766-439-21
; Sequence 21, Application US/08766439
; Patent No. 5922538
;  GENERAL INFORMATION:
;    APPLICANT:  HAZEL, JAMES WILLIAM
;    APPLICANT:  JENSEN, MARK ANTON
;    TITLE OF INVENTION:  GENETIC MARKERS AND METHODS FOR
;    TITLE OF INVENTION:  THE DETECTION OF LISTERIA
;    TITLE OF INVENTION:  MONOCYTOGENES AND LISTERIA SPP.
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.50 INCH DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,439
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/745,228
;      FILING DATE:  NOVEMBER 8, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  MD-1065-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  414 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  YES
;    ORIGINAL SOURCE:
;      STRAIN:  L MONO - 647 - PREMARKER
US-08-766-439-21


  Query Match             50.0%;  Score 22;  DB 2;  Length 414;
  Best Local Similarity   73.7%;  Pred. No. 2.6;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtg 38
          | ||| |||||||||  |  ||||||||  |||  |||
Db    284 AGAAGCATGCGGCCGAGATAATGCGCAACTTATTTGTG 321


RESULT   3
US-08-766-439-28/c
; Sequence 28, Application US/08766439
; Patent No. 5922538
;  GENERAL INFORMATION:
;    APPLICANT:  HAZEL, JAMES WILLIAM
;    APPLICANT:  JENSEN, MARK ANTON
;    TITLE OF INVENTION:  GENETIC MARKERS AND METHODS FOR
;    TITLE OF INVENTION:  THE DETECTION OF LISTERIA
;    TITLE OF INVENTION:  MONOCYTOGENES AND LISTERIA SPP.
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.50 INCH DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,439
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/745,228
;      FILING DATE:  NOVEMBER 8, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  MD-1065-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1274 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      STRAIN:  L MONO 3386 D.F.
US-08-766-439-28


  Query Match             50.0%;  Score 22;  DB 2;  Length 1274;
  Best Local Similarity   73.7%;  Pred. No. 3.2;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtg 38
          | ||| |||||||||  |  ||||||||  |||  |||
Db    501 AGAAGCATGCGGCCGAGATAATGCGCAACTTATTTGTG 464


RESULT   4
US-08-766-439-29
; Sequence 29, Application US/08766439
; Patent No. 5922538
;  GENERAL INFORMATION:
;    APPLICANT:  HAZEL, JAMES WILLIAM
;    APPLICANT:  JENSEN, MARK ANTON
;    TITLE OF INVENTION:  GENETIC MARKERS AND METHODS FOR
;    TITLE OF INVENTION:  THE DETECTION OF LISTERIA
;    TITLE OF INVENTION:  MONOCYTOGENES AND LISTERIA SPP.
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.50 INCH DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
```

```
;       OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;       SOFTWARE:  MICROSOFT WORD 2.0C
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/766,439
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/745,228
;       FILING DATE:  NOVEMBER 8, 1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  FLOYD, LINDA AXAMETHY
;       REGISTRATION NUMBER:  33,692
;       REFERENCE/DOCKET NUMBER:  MD-1065-A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  302-892-8112
;       TELEFAX:  302-773-0164
;   INFORMATION FOR SEQ ID NO:  29:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1274 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ANTI-SENSE:  YES
;     ORIGINAL SOURCE:
;       STRAIN:  L MONO 3386 D.F
US-08-766-439-29


  Query Match             50.0%;  Score 22;  DB 2;  Length 1274;
  Best Local Similarity   73.7%;  Pred. No. 3.2;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgcgcaagatatcagtg 38
           | ||| |||||||||  |  ||||||||  |||  ||| 
Db     774 AGAAGCATGCGGCCGAGATAATGCGCAACTTATTTGTG 811


RESULT   5
US-08-766-439-26/c
; Sequence 26, Application US/08766439
; Patent No. 5922538
;  GENERAL INFORMATION:
;    APPLICANT:  HAZEL, JAMES WILLIAM
;    APPLICANT:  JENSEN, MARK ANTON
;    TITLE OF INVENTION:  GENETIC MARKERS AND METHODS FOR
;    TITLE OF INVENTION:  THE DETECTION OF LISTERIA
;    TITLE OF INVENTION:  MONOCYTOGENES AND LISTERIA SPP.
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.50 INCH DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,439
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/745,228
;      FILING DATE:  NOVEMBER 8, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  MD-1065-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1327 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      STRAIN:  L MONO 899 D.F.
US-08-766-439-26


  Query Match             50.0%;  Score 22;  DB 2;  Length 1327;
  Best Local Similarity   73.7%;  Pred. No. 3.2;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgcgcaagatatcagtg 38
           | ||| |||||||||  |  ||||||||  |||  ||| 
Db     549 AGAAGCATGCGGCCGAGATAATGCGCAACTTATTTGTG 512


RESULT   6
US-08-766-439-27
; Sequence 27, Application US/08766439
; Patent No. 5922538
;  GENERAL INFORMATION:
;    APPLICANT:  HAZEL, JAMES WILLIAM
;    APPLICANT:  JENSEN, MARK ANTON
;    TITLE OF INVENTION:  GENETIC MARKERS AND METHODS FOR
;    TITLE OF INVENTION:  THE DETECTION OF LISTERIA
;    TITLE OF INVENTION:  MONOCYTOGENES AND LISTERIA SPP.
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.50 INCH DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,439
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/745,228
;      FILING DATE:  NOVEMBER 8, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  MD-1065-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  27:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1327 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  YES
;    ORIGINAL SOURCE:
;      STRAIN:  L MONO 899 D.F.
US-08-766-439-27
```

```
  Query Match           50.0%;  Score 22;  DB 2;  Length 1327;
  Best Local Similarity  73.7%;  Pred. No. 3.2;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtg 38
          | ||| |||||||||  |  ||||||||  |||  |||
Db    779 AGAAGCATGCGGCCGAGATAATGCGCAACTTATTTGTG 816


RESULT   7
US-08-766-439-24/c
; Sequence 24, Application US/08766439
; Patent No. 5922538
;  GENERAL INFORMATION:
;    APPLICANT:  HAZEL, JAMES WILLIAM
;    APPLICANT:  JENSEN, MARK ANTON
;    TITLE OF INVENTION:  GENETIC MARKERS AND METHODS FOR
;    TITLE OF INVENTION:  THE DETECTION OF LISTERIA
;    TITLE OF INVENTION:  MONOCYTOGENES AND LISTERIA SPP.
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.50 INCH DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,439
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/745,228
;      FILING DATE:  NOVEMBER 8, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  MD-1065-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  24:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1354 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      STRAIN:  L MONO - 647 - D. FRAG
US-08-766-439-24


  Query Match           50.0%;  Score 22;  DB 2;  Length 1354;
  Best Local Similarity  73.7%;  Pred. No. 3.2;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtg 38
          | ||| |||||||||  |  ||||||||  |||  |||
Db    576 AGAAGCATGCGGCCGAGATAATGCGCAACTTATTTGTG 539


RESULT   8
US-08-766-439-25
; Sequence 25, Application US/08766439
; Patent No. 5922538
;  GENERAL INFORMATION:
;    APPLICANT:  HAZEL, JAMES WILLIAM
;    APPLICANT:  JENSEN, MARK ANTON
;    TITLE OF INVENTION:  GENETIC MARKERS AND METHODS FOR
;    TITLE OF INVENTION:  THE DETECTION OF LISTERIA
;    TITLE OF INVENTION:  MONOCYTOGENES AND LISTERIA SPP.
;    NUMBER OF SEQUENCES:  110
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  U.S.A.
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.50 INCH DISKETTE
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/766,439
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/745,228
;      FILING DATE:  NOVEMBER 8, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA AXAMETHY
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  MD-1065-A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1354 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  YES
;    ORIGINAL SOURCE:
;      STRAIN:  L MONO 647 - D.F.
US-08-766-439-25


  Query Match           50.0%;  Score 22;  DB 2;  Length 1354;
  Best Local Similarity  73.7%;  Pred. No. 3.2;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtg 38
          | ||| |||||||||  |  ||||||||  |||  |||
Db    779 AGAAGCATGCGGCCGAGATAATGCGCAACTTATTTGTG 816


RESULT   9
US-08-974-022-16
; Sequence 16, Application US/08974022
; Patent No. 6015938
;  GENERAL INFORMATION:
;    APPLICANT:  Boyle, Willaim J.
;    APPLICANT:  Lacey, David L.
;    APPLICANT:  Calzone, Frank J.
;    APPLICANT:  Chang, Ming-Shi
;    TITLE OF INVENTION:  OSTEOPROTEGERIN
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Inc.
;      STREET:  1840 Dehavilland Drive
;      CITY:  Thousand Oaks
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91320-1789
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/974,022
;      FILING DATE:  12-DEC-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/577,788
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Winter, Robert B.
;      REFERENCE/DOCKET NUMBER:  A-378
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  45 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-974-022-16


  Query Match           46.8%;  Score 20.6;  DB 3;  Length 45;
  Best Local Similarity  67.4%;  Pred. No. 6.5;
  Matches   29;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtgggaat 43
          |||||||||||||||| |   |  | || |   |||||   ||
Db      1 ATAAGAATGCGGCCGCTAAACTATGAAACAGCCCAGTGACCAT 43


RESULT  10
US-08-795-445A-16
; Sequence 16, Application US/08795445A
; Patent No. 6284485
;  GENERAL INFORMATION:
;    APPLICANT:  Boyle, Willaim J.
;    APPLICANT:  Lacey, David L.
;    APPLICANT:  Calzone, Frank J.
;    APPLICANT:  Chang, Ming-Shi
;    TITLE OF INVENTION:  OSTEOPROTEGERIN
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Inc.
;      STREET:  1840 Dehavilland Drive
;      CITY:  Thousand Oaks
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91320-1789
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/795,445A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/577,788
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Winter, Robert B.
;      REFERENCE/DOCKET NUMBER:  A-378
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  45 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-795-445A-16


  Query Match           46.8%;  Score 20.6;  DB 4;  Length 45;
  Best Local Similarity  67.4%;  Pred. No. 6.5;
  Matches   29;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtgggaat 43
          |||||||||||||||| |   |  | || |   |||||   ||
Db      1 ATAAGAATGCGGCCGCTAAACTATGAAACAGCCCAGTGACCAT 43


RESULT  11
US-08-795-447A-16
; Sequence 16, Application US/08795447A
; Patent No. 6284728
;  GENERAL INFORMATION:
;    APPLICANT:  Boyle, William J.
;    APPLICANT:  Lacey, David L.
;    APPLICANT:  Calzone, Frank J.
;    APPLICANT:  Chang, Ming-Shi
;    TITLE OF INVENTION:  Osteoprotegerin
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Inc.
;      STREET:  One Amgen Center Drive
;      CITY:  Thousand Oaks
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91362-1789
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/795,447A
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Winter, Robert B.
;      REFERENCE/DOCKET NUMBER:  A-378D2
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  45 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-795-447A-16


  Query Match           46.8%;  Score 20.6;  DB 4;  Length 45;
  Best Local Similarity  67.4%;  Pred. No. 6.5;
  Matches   29;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtgggaat 43
          |||||||||||||||| |   |  | || |   |||||   ||
Db      1 ATAAGAATGCGGCCGCTAAACTATGAAACAGCCCAGTGACCAT 43


RESULT  12
US-08-974-186-16
; Sequence 16, Application US/08974186
; Patent No. 6284740
;  GENERAL INFORMATION:
;    APPLICANT:  Boyle, Willaim J.
;    APPLICANT:  Lacey, David L.
;    APPLICANT:  Calzone, Frank J.
;    APPLICANT:  Chang, Ming-Shi
;    TITLE OF INVENTION:  OSTEOPROTEGERIN
;    NUMBER OF SEQUENCES:  53
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Inc.
```

```
;       STREET:  1840 Dehavilland Drive
;       CITY:  Thousand Oaks
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  91320-1789
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/974,186
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/577,788
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Winter, Robert B.
;       REFERENCE/DOCKET NUMBER:  A-378
;   INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  45 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-974-186-16


  Query Match             46.8%;  Score 20.6;  DB 4;  Length 45;
  Best Local Similarity   67.4%;  Pred. No. 6.5;
  Matches   29;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtgggaat 43
          |||||||||||||||| |    |  | || |   |||||   ||
Db      1 ATAAGAATGCGGCCGCTAAACTATGAAACAGCCCAGTGACCAT 43


RESULT  13
US-08-795-446B-16
; Sequence 16, Application US/08795446B
; Patent No. 6288032
;  GENERAL INFORMATION:
;     APPLICANT:  Boyle, Willaim J.
;     APPLICANT:  Lacey, David L.
;     APPLICANT:  Calzone, Frank J.
;     APPLICANT:  Chang, Ming-Shi
;     TITLE OF INVENTION:  OSTEOPROTEGERIN
;     NUMBER OF SEQUENCES:  53
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Amgen Inc.
;       STREET:  1840 Dehavilland Drive
;       CITY:  Thousand Oaks
;       STATE:  California
;       COUNTRY:  USA
;       ZIP:  91320-1789
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/795,446B
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/577,788
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Winter, Robert B.
;       REFERENCE/DOCKET NUMBER:  A-378
;   INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  45 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-795-446B-16


  Query Match             46.8%;  Score 20.6;  DB 4;  Length 45;
  Best Local Similarity   67.4%;  Pred. No. 6.5;
  Matches   29;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtgggaat 43
          |||||||||||||||| |    |  | || |   |||||   ||
Db      1 ATAAGAATGCGGCCGCTAAACTATGAAACAGCCCAGTGACCAT 43


RESULT  14
US-08-158-232-9/c
; Sequence 9, Application US/08158232
; Patent No. 5596071
;  GENERAL INFORMATION:
;     APPLICANT:  Payne, Jewel
;     APPLICANT:  Kennedy, M. Keith
;     APPLICANT:  Randall, John Brooks
;     APPLICANT:  Meier, Henry
;     APPLICANT:  Uick, Heidi Jane
;     APPLICANT:  Foncerrada, Luis
;     APPLICANT:  Schnepf, H. Ernest
;     APPLICANT:  Schwab, George E.
;     APPLICANT:  Fu, Jenny
;     TITLE OF INVENTION:  No. 5596071el Bacillus thuringiensis Toxins Active
;     TITLE OF INVENTION:  Against Hymenopteran Pests
;     NUMBER OF SEQUENCES:  51
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  David R. Saliwanchik
;       STREET:  2421 N.W. 41st Street, Suite A-1
;       CITY:  Gainesville
;       STATE:  FL
;       COUNTRY:  USA
;       ZIP:  32606
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/158,232
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/887,980
;       FILING DATE:  22-MAY-1992
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/797,645
;       FILING DATE:  25-NOV-1991
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/703,977
;       FILING DATE:  22-MAY-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Saliwanchik, David R.
;       REGISTRATION NUMBER:  31,794
;       REFERENCE/DOCKET NUMBER:  M/SCJ104.C1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  904-375-8100
;       TELEFAX:  904-372-5800
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2412 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
```

; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: PS63B
; IMMEDIATE SOURCE:
; CLONE: E. coli NM522(pMYC1642) NRRL B-18961
US-08-158-232-9

Query Match 46.8%; Score 20.6; DB 1; Length 2412;
Best Local Similarity 74.3%; Pred. No. 13;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 9 gcggccgccaccatgcgcaagatatcagtgggaat 43
||| ||||||| || ||||||| |||||||
Db 2268 GCGATCGCCACCCACCGAAAGATATATTTGGGAAT 2234

RESULT 15
US-08-304-626-9/c
; Sequence 9, Application US/08304626
; Patent No. 5616495
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; APPLICANT: Kennedy, M. Keith
; APPLICANT: Randall, John Brooks
; APPLICANT: Meier, Henry
; APPLICANT: Uick, Heidi Jane
; APPLICANT: Foncerrada, Luis
; APPLICANT: Schnepf, Harry E.
; APPLICANT: Schwab, George E.
; TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates
; TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding
; TITLE OF INVENTION: Hymenopteran-Active Toxins
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/304,626
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/887,980
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: M/SCJ 104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2412 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: PS63B
; IMMEDIATE SOURCE:
; CLONE: E. coli NM522(pMYC1642) NRRL B-18961
US-08-304-626-9

Query Match 46.8%; Score 20.6; DB 1; Length 2412;
Best Local Similarity 74.3%; Pred. No. 13;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 9 gcggccgccaccatgcgcaagatatcagtgggaat 43
||| ||||||| || ||||||| |||||||
Db 2268 GCGATCGCCACCCACCGAAAGATATATTTGGGAAT 2234

Search completed: July 31, 2002, 19:32:59
Job time: 8475 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 18:50:09 ; Search time 337.68 Seconds
(without alignments)
223.715 Million cell updates/sec

Title: US-09-824-567-3
Perfect score: 44
Sequence: 1 ataagaatgcggccgccacc.....gcaagatatcagtgggaatc 44

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	44	100.0	44	22	AAD20239	Chlamydia pneumoni
	2	27.6	62.7	49	24	AAS18764	PCR primer #5 used
	3	26.2	59.5	42	24	AAS18778	PCR primer #19 use
c	4	26	59.1	1038602	20	AAZ01425	Complete genome se
	5	25.8	58.6	43	22	AAD09314	Chlamydia pneumoni
	6	25.8	58.6	45	22	AAD20895	C. pneumoniae myos
	7	25.4	57.7	51	21	AAZ56940	C. pneumoniae mip
	8	25.4	57.7	51	21	AAZ56943	C. pneumoniae omp
	9	25	56.8	45	24	AAS18762	PCR primer #3 used
	10	25	56.8	46	22	AAD09149	Chlamydia pneumoni
	11	24.8	56.4	45	21	AAA75902	PCR primer for DNA
	12	24.8	56.4	45	22	AAF83847	C. pneumoniae memb
	13	24.8	56.4	45	24	AAS18774	PCR primer #15 use
	14	24.8	56.4	48	24	AAS18768	PCR primer #9 used
	15	24.6	55.9	43	21	AAZ61587	5' PCR primer for
	16	24.4	55.5	44	22	AAF83843	C. pneumoniae amin
	17	24	54.5	39	21	AAA27123	Chlamydia pneumoni
	18	24	54.5	42	22	AAD20551	C. pneumoniae myos
	19	24	54.5	42	22	AAD20879	C. *pneumoniae myos*
	20	24	54.5	42	22	AAF84487	Chlamydia pneumoni
	21	24	54.5	46	24	AAS18770	PCR primer #11 use
	22	24	54.5	1235	24	AAS18752	Chlamydophila pneu
	23	24	54.5	1799	22	AAD20238	Chlamydia pneumoni
	24	24	54.5	1230025	20	AAX91990	Nucleotide sequenc
	25	23.6	53.6	43	21	AAA30923	PCR primer for C.
	26	23.6	53.6	46	24	AAS18776	PCR primer #17 use
	27	23.6	53.6	2915	23	AAS88477	DNA encoding novel
	28	23.4	53.2	43	21	AAD02069	5' primer for ampl
	29	23.4	53.2	44	22	AAD03025	Chlamydia pneumoni
	30	23.4	53.2	44	24	AAS18766	PCR primer #7 used
	31	23.4	53.2	45	22	AAD20958	C. pneumoniae glut
	32	23.4	53.2	45	22	AAF57426	C. pneumoniae lpdA
	33	23.2	52.7	43	21	AAD02067	5' primer for ampl
	34	23.2	52.7	43	21	AAA75884	PCR primer for DNA
	35	23.2	52.7	43	22	AAD20940	Chlamydia pneumoni
c	36	23.2	52.7	1638	21	AAC39109	Arabidopsis thalia
	37	23	52.3	39	22	AAH46978	Chlamydia general
	38	23	52.3	42	21	AAA28409	5' primer for Chla
	39	23	52.3	42	21	AAA27019	Chlamydia pneumoni
	40	23	52.3	42	22	AAF31255	Chlamydia pneumoni
	41	23	52.3	42	24	AAS18772	PCR primer #13 use
	42	23	52.3	43	21	AAA48840	Chlamydia pneumoni
	43	23	52.3	43	21	AAA28412	5' primer for Chla
	44	23	52.3	43	22	AAF89933	PCR primer used to
	45	23	52.3	43	24	AAS18760	PCR primer #1 used

ALIGNMENTS

RESULT 1
AAD20239
ID AAD20239 standard; DNA; 44 BP.
XX
AC AAD20239;
XX
DT 15-JAN-2002 (first entry)
XX
DE Chlamydia pneumoniae ATP-binding cassette gene amplifying 5'PCR primer.
XX
KW ATP-binding cassette; antibiotic; vaccine; infection; therapy; poxvirus;
KW PCR primer; ss.
XX
OS Chlamydia pneumoniae.
XX
PN WO200174863-A2.
XX
PD 11-OCT-2001.
XX
PF 04-APR-2001; 2001WO-CA00455.
XX
PR 04-APR-2000; 2000US-194464P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI; 2001-648549/74.
XX
PT Novel Chlamydia ATP-binding cassette and corresponding DNA molecule for
PT preventing, diagnosing and treating Chlamydia infections in mammals, in
PT particular humans -

XX
PS Claim 41; Page 53; 88pp; English.
XX
CC The present invention relates to novel Chlamydia pneumoniae ATP-binding
CC cassette protein and its corresponding gene. Sequences of the invention
CC are useful for detecting Chlamydia infection by assaying a body fluid
CC of a mammal with the components. They are also used as vaccines. ATP
CC binding cassette antibodies and vaccines of the invention are useful
CC for preventing or treating Chlamydia infection e.g. infection caused
CC by C. trachomatis, C. psittaci, C. pneumoniae or C. pecorum in mammals,
CC such as humans. The nucleic acid molecules are useful for producing
CC ATP-binding cassettes, in the construction of vaccine vectors such
CC as poxviruses, which are further useful for preventing and/or treating
CC Chlamydia infection and in the construction of attenuated Chlamydia
CC strains that can over-express the nucleic acid molecules or express
CC it in a non-toxic, mutated form. The present DNA sequence is a 5' PCR
CC primer which is used for amplifying Chlamydia pneumoniae ATP-binding
CC cassette DNA.
XX
SQ Sequence 44 BP; 14 A; 11 C; 12 G; 7 T; 0 other;

Query Match 100.0%; Score 44; DB 22; Length 44;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaagatatcagtgggaatc 44
||||||||||||||||||||||||||||||||||||||||||||
Db 1 ataagaatgcggccgccaccatgcgcaagatatcagtgggaatc 44

RESULT 2
AAS18764
ID AAS18764 standard; DNA; 49 BP.
XX
AC AAS18764;
XX
DT 26-MAR-2002 (first entry)
XX
DE PCR primer #5 used to amplify Chlamydophila pneumoniae gene.
XX
KW ATP binding cassette; secretary locus open reading frame; endopeptidase;
KW secretary locus ORF; protease; metalloprotease; CLP protease ATPase;
KW CLP protease subunit; transglycolase/transpeptidase; CLPc protease;
KW thioredoxin; Chlamydia infection; antibacterial; PCR primer; ss.
XX
OS Chlamydophila pneumoniae CWL029.
XX
PN WO200185972-A2.
XX
PD 15-NOV-2001.
XX
PF 08-MAY-2001; 2001WO-CA00653.
XX
PR 08-MAY-2000; 2000US-202672P.
PR 30-MAY-2000; 2000US-207852P.
PR 16-JUN-2000; 2000US-211796P.
PR 16-JUN-2000; 2000US-211797P.
PR 16-JUN-2000; 2000US-211798P.
PR 16-JUN-2000; 2000US-211801P.
PR 16-JUN-2000; 2000US-212044P.
PR 26-SEP-2000; 2000US-235335P.
PR 26-SEP-2000; 2000US-235361P.
PR 26-SEP-2000; 2000US-235398P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI; 2002-049447/06.
XX
PT Vaccine useful for immunising mammals against chlamydia infections,
PT comprises vectors having sequences of ATP binding cassette gene,
PT secretary locus open reading frame gene of chlamydia -
XX
PS Example 1; Page 63; 355pp; English.
XX
CC The present invention relates to the isolation of Chlamydophila
CC pneumoniae strain CWL029 genes and their encoded proteins. The genes of
CC the invention encode an ATP binding cassette gene, a secretary locus
CC open reading frame (ORF), an endopeptidase, a protease, a
CC metalloprotease, CLP protease ATPase, a CLP protease subunit, a
CC transglycolase/transpeptidase, a CLPc protease, or thioredoxin. The
CC genes of the invention can be used in a vector as a vaccine for the
CC prevention and treatment of Chlamydia infections. Also described are
CC B- and T-cell epitopes from the proteins of the invention which can be
CC used as Chlamydia antigens. AAS18760-AAS18779 represent PCR primers
CC used to amplify the C. pneumoniae genes (AAS18750-AAS18759) of the
CC invention.
XX
SQ Sequence 49 BP; 12 A; 14 C; 12 G; 11 T; 0 other;

Query Match 62.7%; Score 27.6; DB 24; Length 49;
Best Local Similarity 78.6%; Pred. No. 0.055;
Matches 33; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaagatatcagtgggaa 42
|||||||||||||||||||||||| || || ||| ||
Db 1 ataagaatgcggccgccaccatgctcaccctaggcttggaaa 42

RESULT 3
AAS18778
ID AAS18778 standard; DNA; 42 BP.
XX
AC AAS18778;
XX
DT 26-MAR-2002 (first entry)
XX
DE PCR primer #19 used to amplify Chlamydophila pneumoniae gene.
XX
KW ATP binding cassette; secretary locus open reading frame; endopeptidase;
KW secretary locus ORF; protease; metalloprotease; CLP protease ATPase;
KW CLP protease subunit; transglycolase/transpeptidase; CLPc protease;
KW thioredoxin; Chlamydia infection; antibacterial; PCR primer; ss.
XX
OS Chlamydophila pneumoniae CWL029.
XX
PN WO200185972-A2.
XX
PD 15-NOV-2001.
XX
PF 08-MAY-2001; 2001WO-CA00653.
XX
PR 08-MAY-2000; 2000US-202672P.
PR 30-MAY-2000; 2000US-207852P.
PR 16-JUN-2000; 2000US-211796P.
PR 16-JUN-2000; 2000US-211797P.
PR 16-JUN-2000; 2000US-211798P.
PR 16-JUN-2000; 2000US-211801P.
PR 16-JUN-2000; 2000US-212044P.
PR 26-SEP-2000; 2000US-235335P.
PR 26-SEP-2000; 2000US-235361P.
PR 26-SEP-2000; 2000US-235398P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI; 2002-049447/06.
XX
PT Vaccine useful for immunising mammals against chlamydia infections,
PT comprises vectors having sequences of ATP binding cassette gene,
PT secretary locus open reading frame gene of chlamydia -
XX

PS Example 1; Page 68; 355pp; English.
XX
CC The present invention relates to the isolation of Chlamydophila
CC pneumoniae strain CWL029 genes and their encoded proteins. The genes of
CC the invention encode an ATP binding cassette gene, a secretary locus
CC open reading frame (ORF), an endopeptidase, a protease, a
CC metalloprotease, CLP protease ATPase, a CLP protease subunit, a
CC transglycolase/transpeptidase, a CLPc protease, or thioredoxin. The
CC genes of the invention can be used in a vector as a vaccine for the
CC prevention and treatment of Chlamydia infections. Also described are
CC B- and T-cell epitopes from the proteins of the invention which can be
CC used as Chlamydia antigens. AAS18760-AAS18779 represent PCR primers
CC used to amplify the C. pneumoniae genes (AAS18750-AAS18759) of the
CC invention.
XX
SQ Sequence 42 BP; 15 A; 9 C; 10 G; 8 T; 0 other;

Query Match 59.5%; Score 26.2; DB 24; Length 42;
Best Local Similarity 90.3%; Pred. No. 0.2;
Matches 28; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaagat 31
||||||||||||||||||||||| |||||
Db 1 ataagaatgcggccgccaccatggtaaagat 31

RESULT 4
AAZ01425/c
ID AAZ01425 standard; DNA; 1038602 BP.
XX
AC AAZ01425;
XX
DT 07-OCT-1999 (first entry)
XX
DE Complete genome sequence of Chlamydia trachomatis.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX
OS Chlamydia trachomatis.
XX
PN WO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Claim 1; Page 373-656; 1755pp; English.
XX
CC The present sequence represents the complete genome of Chlamydia
CC trachomatis. Open reading frames (ORFs) of the genome encode
CC polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines
CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
CC be used to control growth of the microorganism. Chlamydia trachomatis is
CC responsible for a large number of diseases, e.g. eye diseases such as
CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
CC conjunctivitis; genital diseases such as nongonococcal uretritis,
CC epidymitis, cervicitis, salpingitis, perihepatitis, bartholinitis;
CC pneumopathy in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
CC treating these diseases.
XX
SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 59.1%; Score 26; DB 20; Length 1038602;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 ccatgcgcaagatatcagtgggaatc 44
||||||||||||||||||||||||||
Db 252099 CCATGCGCAAGATATCAGTGGGAATC 252074

RESULT 5
AAD09314
ID AAD09314 standard; DNA; 43 BP.
XX
AC AAD09314;
XX
DT 10-SEP-2001 (first entry)
XX
DE Chlamydia pneumoniae outer membrane protein gene amplifying 5'PCR primer.
XX
KW Outer membrane protein; therapy; Chlamydia infection;
KW antibiotic; vaccine; PCR primer; ss.
XX
OS Chlamydia pneumoniae.
XX
PN WO200146225-A2.
XX
PD 28-JUN-2001.
XX
PF 20-DEC-2000; 2000WO-CA01535.
XX
PR 22-DEC-1999; 99US-0171539.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI; 2001-418020/44.
XX
PT Chlamydia outer membrane protein and corresponding DNA molecules for
PT preventing, diagnosing and treating Chlamydia infection in mammals,
PT such as humans -
XX
PS Claim 32; Page 52; 74pp; English.
XX
CC The present sequence is a PCR primer which is used to amplify the
CC Chlamydia pneumoniae outer membrane protein gene. The outer membrane
CC protein is useful for preventing, treating and detecting Chlamydia
CC infection in humans. The outer membrane protein DNA is useful for
CC producing the encoded polypeptide and in the construction of attenuated
CC Chlamydia strains that can over express the polynucleotide or express
CC it in a non-toxic, mutated form. It is also used as vaccine. The probes
CC for outer membrane protein are useful in diagnostic tests as capture or
CC detection probes and the primers are useful in diagnostic methods
CC involving PCR. The antibody against outer membrane protein is useful for
CC purifying the outer membrane protein.
XX
SQ Sequence 43 BP; 16 A; 9 C; 6 G; 12 T; 0 other;

Query Match 58.6%; Score 25.8; DB 22; Length 43;
Best Local Similarity 81.1%; Pred. No. 0.29;
Matches 30; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaagatatcagt 37
||||||||||||||||||||||| || ||| | |
Db 1 ataagaatgcggccgccaccatgaaaaaattattatt 37

```
RESULT   6
AAD20895
ID   AAD20895 standard; DNA; 45 BP.
XX
AC   AAD20895;
XX
DT   15-JAN-2002  (first entry)
XX
DE   C. pneumoniae myosin gene amplifying 5' PCR primer alternative version.
XX
KW   Transmembrane protein; antibacterial; vaccine; gene therapy;
KW   immunisation; PCR primer; ss.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO200175114-A2.
XX
PD   11-OCT-2001.
XX
PF   04-APR-2001; 2001WO-CA00462.
XX
PR   04-APR-2000; 2000US-194477P.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
XX
PI   Murdin AD,  Oomen RP,  Wang J,  Dunn P;
XX
DR   WPI; 2001-648559/74.
XX
PT   Novel polypeptides from Chlamydia pneumoniae and genes encoding the
PT   polypeptide, useful for immunization of host e.g. human against disease
PT   caused by infection by a strain of Chlamydia  -
XX
PS   Claim 41; Page 89; 90pp; English.
XX
CC   The invention relates to a transmembrane polypeptide from Chlamydia,
CC   preferably C. pneumoniae. Transmembrane protein and its gene are useful
CC   as vaccines and for preventing or treating Chlamydia infection.
CC   Transmembrane protein, its gene and antibody are useful for detecting
CC   Chlamydia infection, by assaying a body fluid of a mammal to be tested
CC   The probes are used in diagnostic tests as capture or detection probes
CC   and in hybridisation techniques, and primers are useful in amplification
CC   techniques for use in diagnostic methods. Transmembrane protein is useful
CC   for detecting the presence of anti-Chlamydia antibodies in blood sample.
CC   The present sequence is an alternative version of a PCR primer used for
CC   amplifying C. pneumoniae myosin heavy chain homologue gene used in the
CC   exemplification of the invention.
CC   Note: This sequence is stated as being the same as SEQ ID NO: 3 shown
CC   in page 53 (AAD20879) of the specification. However, the sequences differ
CC   at several positions.
XX
SQ   Sequence 45 BP; 12 A; 14 C; 8 G; 11 T; 0 other;


  Query Match            58.6%;  Score 25.8;  DB 22;  Length 45;
  Best Local Similarity  93.1%;  Pred. No. 0.3;
  Matches   27;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaag 29
          ||||||||||||||||||||||||| |||
Db      1 ataagaatgcggccgccaccatgcagaag 29


RESULT   7
AAZ56940
ID   AAZ56940 standard; DNA; 51 BP.
XX
AC   AAZ56940;
XX
DT   08-MAY-2000  (first entry)
XX
DE   C. pneumoniae mip gene amplifying 5' primer.
XX
KW   Chlamydia pneumoniae; outer membrane protein; mip; CPN100501;
KW   chlamydial infection; PCR primer; ss.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO200006741-A1.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-IB01330.
XX
PR   27-JUL-1998;   98US-0094192.
PR   01-MAR-1999;   99US-0122044.
PR   26-JUL-1999;   99US-0361440.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Murdin AD,  Oomen RP,  Dunn PL;
XX
DR   WPI; 2000-195302/17.
XX
PT   Novel polynucleotides and Chlamydia pneumoniae outer membrane protein
PT   encoded by them for use as vaccines in treating and diagnosing
PT   chlamydial infections  -
XX
PS   Example 1; Page 34; 55pp; English.
XX
CC   The invention provides an isolated polynucleotide encoding Chlamydia
CC   pneumoniae outer membrane protein (mip or CPN100501). The mip protein
CC   can be expressed by standard recombinant methodology. the mip gene is
CC   used for detecting Chlamydia by hybridizing or amplifying the sample
CC   with the mip gene specific probe. A vaccine vector or a pharmaceutical
CC   composition comprising the mip gene are used for inducing an immune
CC   response in a mammal to prevent/treat chlamydial infections particularly
CC   infections caused by C. pneumoniae. The present sequence represents the
CC   a PCR primer amplifying the C. pneumoniae mip gene.
XX
SQ   Sequence 51 BP; 20 A; 9 C; 9 G; 13 T; 0 other;


  Query Match            57.7%;  Score 25.4;  DB 21;  Length 51;
  Best Local Similarity  82.9%;  Pred. No. 0.44;
  Matches   29;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatca 35
          |||||||||||||||||||||||   || | || |
Db      1 ataagaatgcggccgccaccatgaaaaaaaaatta 35


RESULT   8
AAZ56943
ID   AAZ56943 standard; DNA; 51 BP.
XX
AC   AAZ56943;
XX
DT   08-MAY-2000  (first entry)
XX
DE   C. pneumoniae omp gene amplifying 5' primer.
XX
KW   Chlamydia pneumoniae antigen; omp; CPN100314; antibacterial; PCR primer;
KW   vaccination; Chlamydia infection; community acquired pneumonia;
KW   upper respiratory tract infection; bronchitis; sinusitis; ss.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO200006743-A2.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-IB01333.
XX
```

```
PR   27-JUL-1998;   98US-0094203.
PR   01-MAR-1999;   99US-0122045.
PR   26-JUL-1999;   99US-0360434.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Murdin AD,  Oomen RP,  Dunn PL;
XX
DR   WPI; 2000-195303/17.
XX
PT   Chlamydia pneumoniae antigens used for immunization and protection
PT   against Chlamydia diseases  -
XX
PS   Example 1; Page 34; 52pp; English.
XX
CC   The invention relates to Chlamydia pneumoniae antigens, and their
CC   corresponding polynucleotides. These omp (CPN100314) polypeptides are
CC   found in the bacterial membrane structure and its external vicinity, in
CC   the inclusion membrane and its external vicinity, and are released into
CC   the cytoplasm of the infected cell. The C. pneumoniae polynucleotides and
CC   polypeptides can be used in vaccination methods for preventing and
CC   treating Chlamydia infection (e.g. infections caused by C. trachomatis,
CC   C. psittaci, C. pneumoniae or C. pecorum). The polynucleotides can be
CC   used to produce the polypeptides recombinantly, in the construction of
CC   vaccine vectors, as a vaccine agent, and in the construction of an
CC   attenuated Chlamydia strain. The polypeptides are also useful as vaccine
CC   agents, and for the preparation of medicaments for treating or preventing
CC   Chlamydia infection, e.g. community acquired pneumonia, and upper
CC   respiratory tract infections such as bronchitis and sinusitis. The
CC   present sequence represents a PCR primer for amplifying C. pneumoniae
CC   omp gene.
xx
SQ   Sequence 51 BP; 20 A; 9 C; 9 G; 13 T; 0 other;
```

```
  Query Match            57.7%;  Score 25.4;  DB 21;  Length 51;
  Best Local Similarity  82.9%;  Pred. No. 0.44;
  Matches   29;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgcgcaagatatca 35
           ||||||||||||||||||||||||   || | || |
Db       1 ataagaatgcggccgccaccatgaaaaaaaaatta 35
```

```
RESULT   9
AAS18762
ID   AAS18762 standard; DNA; 45 BP.
XX
AC   AAS18762;
XX
DT   26-MAR-2002  (first entry)
XX
DE   PCR primer #3 used to amplify Chlamydophila pneumoniae gene.
XX
KW   ATP binding cassette; secretary locus open reading frame; endopeptidase;
KW   secretary locus ORF; protease; metalloprotease; CLP protease ATPase;
KW   CLP protease subunit; transglycolase/transpeptidase; CLPc protease;
KW   thioredoxin; Chlamydia infection; antibacterial; PCR primer; ss.
XX
OS   Chlamydophila pneumoniae CWL029.
XX
PN   WO200185972-A2.
XX
PD   15-NOV-2001.
XX
PF   08-MAY-2001; 2001WO-CA00653.
XX
PR   08-MAY-2000; 2000US-202672P.
PR   30-MAY-2000; 2000US-207852P.
PR   16-JUN-2000; 2000US-211796P.
PR   16-JUN-2000; 2000US-211797P.
PR   16-JUN-2000; 2000US-211798P.
PR   16-JUN-2000; 2000US-211801P.
PR   16-JUN-2000; 2000US-212044P.
PR   26-SEP-2000; 2000US-235335P.
PR   26-SEP-2000; 2000US-235361P.
PR   26-SEP-2000; 2000US-235398P.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
XX
PI   Murdin AD,  Oomen RP,  Wang J,  Dunn P;
XX
DR   WPI; 2002-049447/06.
XX
PT   Vaccine useful for immunising mammals against chlamydia infections,
PT   comprises vectors having sequences of ATP binding cassette gene,
PT   secretary locus open reading frame gene of chlamydia  -
XX
PS   Example 1; Page 62; 355pp; English.
XX
CC   The present invention relates to the isolation of Chlamydophila
CC   pneumoniae strain CWL029 genes and their encoded proteins. The genes of
CC   the invention encode an ATP binding cassette gene, a secretary locus
CC   open reading frame (ORF), an endopeptidase, a protease, a
CC   metalloprotease, CLP protease ATPase, a CLP protease subunit, a
CC   transglycolase/transpeptidase, a CLPc protease, or thioredoxin. The
CC   genes of the invention can be used in a vector as a vaccine for the
CC   prevention and treatment of Chlamydia infections. Also described are
CC   B- and T-cell epitopes from the proteins of the invention which can be
CC   used as Chlamydia antigens. AAS18760-AAS18779 represent PCR primers
CC   used to amplify the C. pneumoniae genes (AAS18750-AAS18759) of the
CC   invention.
XX
SQ   Sequence 45 BP; 13 A; 11 C; 13 G; 8 T; 0 other;
```

```
  Query Match            56.8%;  Score 25;  DB 24;  Length 45;
  Best Local Similarity  100.0%;  Pred. No. 0.63;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgcg 25
           |||||||||||||||||||||||||
Db       1 ataagaatgcggccgccaccatgcg 25
```

```
RESULT  10
AAD09149
ID   AAD09149 standard; DNA; 46 BP.
XX
AC   AAD09149;
XX
DT   10-SEP-2001  (first entry)
XX
DE   Chlamydia pneumoniae membrane ATPase gene amplifying 5' PCR primer.
XX
KW   Membrane ATPase; vaccine; therapy; Chlamydia infection; pneumoniae;
KW   bronchitis; sinusitis; acute respiratory disease; cough; fever; asthma;
KW   abnormal chest sound; lower respiratory tract infection; atherosclerosis;
KW   antibiotic; PCR primer; ss.
XX
OS   Chlamydia pneumoniae.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    21..46
FT                   /*tag=  a
FT                   /note= "5' end of the membrane ATPase coding
FT                   sequence"
XX
PN   WO200146226-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-CA01536.
XX
PR   22-DEC-1999;   99US-0171538.
XX
```

```
PA   (AVET ) AVENTIS PASTEUR LTD.
XX
PI   Murdin AD,  Oomen RP,  Wang J,  Dunn P;
XX
DR   WPI; 2001-418021/44.
XX
PT   Chlamydia polypeptides, designated membrane ATPase and corresponding
PT   DNA molecules for preventing, diagnosing and treating Chlamydia
PT   infection in mammals, including humans -
XX
PS   Claim 32; Page 52; 80pp; English.
XX
CC   The present sequence is a 5' PCR primer used to amplify
CC   Chlamydia pneumoniae membrane ATPase gene. Membrane ATPase is used
CC   as vaccine. Membrane ATPase is useful for detecting, preventing and
CC   treating Chlamydia infection such as pneumoniae, bronchitis, sinusitis,
CC   acute respiratory disease, cough, fever, abnormal chest sounds, lower
CC   respiratory tract infection, atherosclerosis and asthma, in mammals,
CC   in particular humans. Membrane ATPase is useful in the construction of
CC   attenuated Chlamydia strains that can over express the polynucleotide or
CC   express it in a non-toxic, mutated form.
XX
SQ   Sequence 46 BP; 21 A; 13 C; 6 G; 6 T; 0 other;


  Query Match             56.8%;  Score 25;  DB 22;  Length 46;
  Best Local Similarity   84.8%;  Pred. No. 0.63;
  Matches   28;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgcgcaagatat 33
           |||||||||||||||||||||||||  |  || |
Db       1 ataagaatgcggccgccaccatgcaaacaatct 33


RESULT  11
AAA75902
ID   AAA75902 standard; DNA; 45 BP.
XX
AC   AAA75902;
XX
DT   22-JAN-2001  (first entry)
XX
DE   PCR primer for DNA encoding a 60 kDa cysteine-rich membrane protein.
XX
KW   Cysteine-rich membrane protein; Chlamydia infection; bronchitis;
KW   community acquired pneumonia; upper respiratory tract infection; vaccine;
KW   sinusitis; PCR primer; ss.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO200055326-A1.
XX
PD   21-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-CA00240.
XX
PR   12-MAR-1999;   99US-0123966.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
XX
PI   Murdin AD,  Oomen RP,  Wang J,  Dunn P;
XX
DR   WPI; 2000-618918/59.
XX
PT   New polynucleotides encoding a 60kda cysteine-rich membrane protein
PT   from Chlamydia, useful as a vaccine for preventing and treating
PT   Chlamydia infection in mammals  -
XX
PS   Example 1; Page 48; 77pp; English.
XX
CC   PCR primers AAA75902-03 were used to amplify DNA encoding a Chlamydia
CC   60 kDa cysteine-rich membrane protein. The membrane-rich polynucleotide
CC   and polypeptide are useful for preventing or treating Chlamydia
CC   infection, such as community pneumonia, upper respiratory tract
CC   infections, bronchitis and sinusitis. They are also useful for
CC   diagnosing Chlamydia infection by assaying a body fluid of a mammal.
CC   The polypeptide is useful for vaccine production.
XX
SQ   Sequence 45 BP; 15 A; 13 C; 10 G; 7 T; 0 other;


  Query Match             56.4%;  Score 24.8;  DB 21;  Length 45;
  Best Local Similarity   92.9%;  Pred. No. 0.76;
  Matches   26;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgcgcaa 28
           ||||||||||||||||||||||||| |||
Db       1 ataagaatgcggccgccaccatgtccaa 28


RESULT  12
AAF83847
ID   AAF83847 standard; DNA; 45 BP.
XX
AC   AAF83847;
XX
DT   06-AUG-2001  (first entry)
XX
DE   C. pneumoniae membrane ATPase DNA amplifying 5' primer.
XX
KW   Membrane adenosine triphosphatase; ATP; membrane ATPase; Chlamydia;
KW   infection; antibiotic; vaccine; PCR primer; ss.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO200136455-A2.
XX
PD   25-MAY-2001.
XX
PF   10-NOV-2000; 2000WO-CA01344.
XX
PR   12-NOV-1999;   99US-0164823.
XX
PA   (AVET ) AVENTIS PASTEUR LTD.
XX
PI   Murdin AD,  Oomen RP,  Wang J,  Dunn P;
XX
DR   WPI; 2001-355610/37.
XX
PT   Novel Chlamydia pneumoniae membrane ATPase polypeptide useful for
PT   preventing, treating or detecting Chlamydia infection -
XX
PS   Claim 32; Page 52; 83pp; English.
XX
CC   Sequences AAF83847-848 represent PCR primers for amplifying the genomic
CC   DNA encoding a Chlamydia pneumoniae membrane adenosine triphosphate
CC   (ATP)ase polypeptide. The membrane ATPase polynucleotides and polypeptide
CC   and fragments are useful for preventing or treating Chlamydia infection
CC   and for detecting Chlamydia by assaying a body fluid. The polynucleotides
CC   are useful for producing recombinant membrane ATPase, in the construction
CC   of vaccine vectors for preventing and/or treating Chlamydia infection,
CC   as a vaccine agent, and in the construction of attenuated Chlamydia
CC   strains that can over-express the polypeptide, or expresses it in a
CC   non-toxic, mutated form.
XX
SQ   Sequence 45 BP; 16 A; 11 C; 10 G; 8 T; 0 other;


  Query Match             56.4%;  Score 24.8;  DB 22;  Length 45;
  Best Local Similarity   80.6%;  Pred. No. 0.76;
  Matches   29;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgcgcaagatatcag 36
           |||||||||||||||||||||||   |   | ||||
Db       1 ataagaatgcggccgccaccatggtaacagtttcag 36
```

RESULT 13
AAS18774
ID AAS18774 standard; DNA; 45 BP.
XX
AC AAS18774;
XX
DT 26-MAR-2002 (first entry)
XX
DE PCR primer #15 used to amplify Chlamydophila pneumoniae gene.
XX
KW ATP binding cassette; secretary locus open reading frame; endopeptidase;
KW secretary locus ORF; protease; metalloprotease; CLP protease ATPase;
KW CLP protease subunit; transglycolase/transpeptidase; CLPc protease;
KW thioredoxin; Chlamydia infection; antibacterial; PCR primer; ss.
XX
OS Chlamydophila pneumoniae CWL029.
XX
PN WO200185972-A2.
XX
PD 15-NOV-2001.
XX
PF 08-MAY-2001; 2001WO-CA00653.
XX
PR 08-MAY-2000; 2000US-202672P.
PR 30-MAY-2000; 2000US-207852P.
PR 16-JUN-2000; 2000US-211796P.
PR 16-JUN-2000; 2000US-211797P.
PR 16-JUN-2000; 2000US-211798P.
PR 16-JUN-2000; 2000US-211801P.
PR 16-JUN-2000; 2000US-212044P.
PR 26-SEP-2000; 2000US-235335P.
PR 26-SEP-2000; 2000US-235361P.
PR 26-SEP-2000; 2000US-235398P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI; 2002-049447/06.
XX
PT Vaccine useful for immunising mammals against chlamydia infections,
PT comprises vectors having sequences of ATP binding cassette gene,
PT secretary locus open reading frame gene of chlamydia -
XX
PS Example 1; Page 66; 355pp; English.
XX
CC The present invention relates to the isolation of Chlamydophila
CC pneumoniae strain CWL029 genes and their encoded proteins. The genes of
CC the invention encode an ATP binding cassette gene, a secretary locus
CC open reading frame (ORF), an endopeptidase, a protease, a
CC metalloprotease, CLP protease ATPase, a CLP protease subunit, a
CC transglycolase/transpeptidase, a CLPc protease, or thioredoxin. The
CC genes of the invention can be used in a vector as a vaccine for the
CC prevention and treatment of Chlamydia infections. Also described are
CC B- and T-cell epitopes from the proteins of the invention which can be
CC used as Chlamydia antigens. AAS18760-AAS18779 represent PCR primers
CC used to amplify the C. pneumoniae genes (AAS18750-AAS18759) of the
CC invention.
XX
SQ Sequence 45 BP; 13 A; 14 C; 10 G; 8 T; 0 other;

Query Match 56.4%; Score 24.8; DB 24; Length 45;
Best Local Similarity 92.9%; Pred. No. 0.76;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaa 28
||||||||||||||||||||||| || |
Db 1 ataagaatgcggccgccaccatgagcta 28

RESULT 14
AAS18768
ID AAS18768 standard; DNA; 48 BP.
XX
AC AAS18768;
XX
DT 26-MAR-2002 (first entry)
XX
DE PCR primer #9 used to amplify Chlamydophila pneumoniae gene.
XX
KW ATP binding cassette; secretary locus open reading frame; endopeptidase;
KW secretary locus ORF; protease; metalloprotease; CLP protease ATPase;
KW CLP protease subunit; transglycolase/transpeptidase; CLPc protease;
KW thioredoxin; Chlamydia infection; antibacterial; PCR primer; ss.
XX
OS Chlamydophila pneumoniae CWL029.
XX
PN WO200185972-A2.
XX
PD 15-NOV-2001.
XX
PF 08-MAY-2001; 2001WO-CA00653.
XX
PR 08-MAY-2000; 2000US-202672P.
PR 30-MAY-2000; 2000US-207852P.
PR 16-JUN-2000; 2000US-211796P.
PR 16-JUN-2000; 2000US-211797P.
PR 16-JUN-2000; 2000US-211798P.
PR 16-JUN-2000; 2000US-211801P.
PR 16-JUN-2000; 2000US-212044P.
PR 26-SEP-2000; 2000US-235335P.
PR 26-SEP-2000; 2000US-235361P.
PR 26-SEP-2000; 2000US-235398P.
XX
PA (AVET) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
DR WPI; 2002-049447/06.
XX
PT Vaccine useful for immunising mammals against chlamydia infections,
PT comprises vectors having sequences of ATP binding cassette gene,
PT secretary locus open reading frame gene of chlamydia -
XX
PS Example 1; Page 64; 355pp; English.
XX
CC The present invention relates to the isolation of Chlamydophila
CC pneumoniae strain CWL029 genes and their encoded proteins. The genes of
CC the invention encode an ATP binding cassette gene, a secretary locus
CC open reading frame (ORF), an endopeptidase, a protease, a
CC metalloprotease, CLP protease ATPase, a CLP protease subunit, a
CC transglycolase/transpeptidase, a CLPc protease, or thioredoxin. The
CC genes of the invention can be used in a vector as a vaccine for the
CC prevention and treatment of Chlamydia infections. Also described are
CC B- and T-cell epitopes from the proteins of the invention which can be
CC used as Chlamydia antigens. AAS18760-AAS18779 represent PCR primers
CC used to amplify the C. pneumoniae genes (AAS18750-AAS18759) of the
CC invention.
XX
SQ Sequence 48 BP; 17 A; 11 C; 8 G; 12 T; 0 other;

Query Match 56.4%; Score 24.8; DB 24; Length 48;
Best Local Similarity 92.9%; Pred. No. 0.77;
Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ataagaatgcggccgccaccatgcgcaa 28
||||||||||||||||||||||| | ||
Db 1 ataagaatgcggccgccaccatgagaaa 28

RESULT 15
AAZ61587
ID AAZ61587 standard; DNA; 43 BP.

XX
AC AAZ61587;
XX
DT 19-JUN-2000 (first entry)
XX
DE 5' PCR primer for POMP91A gene of Chlamydia pneumoniae.
XX
KW POMP91A; Chlamydia pneumoniae strain CMI; Chlamydia infection;
KW vaccine; immune response; PCR primer; ss.
XX
OS Chlamydia pneumoniae.
XX
PN WO200011180-A1.
XX
PD 02-MAR-2000.
XX
PF 19-AUG-1999; 99WO-CA00765.
XX
PR 20-AUG-1998; 98US-0097198.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Murdin AD, Dunn PL, Oomen RP;
XX
DR WPI; 2000-224700/19.
XX
PT New nucleic acid encoding POMP91A protein from a strain of Chlamydia
PT useful for preventing, treating and diagnosing Chlamydia infection -
XX
PS Example; Page 38; 98pp; English.
XX
CC PCR primers AAZ61587-88 were used to amplify DNA encoding a polypeptide
CC of POMP91A from Chlamydia pneumoniae strain CMI genomic DNA. The
CC polynucleotides or polypeptides are used to prevent, treat and
CC diagnose Chlamydia infection. Vaccine vectors containing POMP91A
CC polynucleotides are used to induce an immune response against
CC Chlamydia. Antibodies against POMP91A can be used to diagnose the
CC presence of Chlamydia in a biological sample.
XX
SQ Sequence 43 BP; 11 A; 10 C; 14 G; 8 T; 0 other;

Query Match 55.9%; Score 24.6; DB 21; Length 43;
Best Local Similarity 87.1%; Pred. No. 0.91;
Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 ataagaatgcggccgccaccatgcgcaagat 31
          ||||||||||| |||||||||||    ||||
Db      1 ataagaatgcggccgccaccatgaagcagat 31
```

Search completed: July 31, 2002, 20:58:36
Job time: 7707 sec

```
OM of: US-09-824-567-2 to: EST:*   out_format : pfs

Date: Jul 26, 2002  5:12 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09824567/runat_23072002_141350_13438/app_query.fasta_1.596
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US09824567_@CGN1_1_5096
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-824-567-2
Query length: 532
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1769.610000


score_list:
Sequence                Strd Orig    ZScore    EScore Len  ! Documentation  ..
gb_est1:AI777353          -   167.50  300.79   1.7e-07   644  ! AI777353 EST263761 tomato seed,
gb_gss:AQ990639           +   148.50  266.40   1.4e-05   538  ! AQ990639 Rfc01442 Photorhabdus
gb_gss:BH370951           -   137.00  243.09   0.0003    593  ! BH370951 AG-ND-119H23.TF ND-TAM
gb_gss:B07758             -   124.50  218.87   0.0061    601  ! B07758 8219PI01E01120195T7 Rhod
gb_gss:AQ990866           -   122.50  214.78   0.0102    613  ! AQ990866 Rfc01703 Photorhabdus
gb_gss:AZ302752           +   118.00  211.22   0.0162    402  ! AZ302752 GSSBru1690 Brucella ab
gb_gss:AQ012177           +   114.50  197.56   0.0932    713  ! AQ012177 27I1C073112697 Cosmid
gb_est1:AI503668          +   109.00  194.26   0.1421    390  ! AI503668 vk75e04.x1 Knowles Sol
gb_gss:AQ579158           +   108.50  187.24   0.3499    644  ! AQ579158 nbxb0096M13f CUGI Rice
gb_gss:B07745             -   104.00  181.68   0.7133    498  ! B07745 8025RI01A070828955T3 Rhod
gb_gss:TA380G06P          +   103.50  180.22   0.8603    519  ! AL497735 T. brucei sheared genc
gb_gss:BH398784           -   103.50  179.43   0.9519    554  ! BH398784 AG-ND-150B19.TR ND-TAM
gb_gss:TA358F03P          -   103.50  179.35   0.9626    558  ! AL494115 T. brucei sheared genc
gb_gss:AZ302822           +   102.50  181.72   0.7099    391  ! AZ302822 GSSBru1761 Brucella ab
gb_gss:BH375296           -   102.50  176.89   1.32      583  ! BH375296 AG-ND-174G16.TR ND-TAM
gb_est2:BF574618          -   102.00  169.88   3.24      962  ! BF574618 602133277F1 NIH_MGC_81
gb_est2:BI180815          -   101.50  179.98   0.8872    385  ! BI180815 3L21C3 LRH (Lin Rhodam
gb_est2:BI180690          +   101.50  179.95   0.8908    386  ! BI180690 LRH21C3 LRH (Lin Rhoda
gb_est2:BI180755          +   101.50  179.89   0.8979    388  ! BI180755 LRH24F12 LRH (Lin Rhod
gb_est2:BF972232          +   101.50  163.69   7.17      1482  ! BF972232 602240703F1 NIH_MGC_4
gb_est2:BE793898          +   100.50  165.46   5.72      1092  ! BE793898 601592727F1 NIH_MGC_7
gb_est2:BJ032136          +    99.00  169.95   3.21      593  ! BJ032136 BJ032136 NIBB Mochii m
gb_est2:BF534669          +    99.00  165.40   5.76      864  ! BF534669 602048193F1 NCI_CGAP_S
gb_gss:CNS07897           +    98.50  161.72   9.23      1082  ! AL433697 T7 end of clone BB0AA
gb_gss:CNS04XA3           -    98.00  162.47   8.38      939  ! AL311412 Tetraodon nigroviridis
gb_est1:AA549415          +    97.50  173.89   1.94      337  ! AA549415 vk75e04.s1 Knowles Sol
gb_gss:AF081360           +    97.50  169.17   3.55      498  ! AF081360 AF081360 Thermus therm
gb_est2:BF663011          +    97.00  155.34   20.93     1445  ! BF663011 602144913F1 NIH_MGC_4
gb_est2:BE909270          +    96.50  157.99   14.89     1071  ! BE909270 601502920F1 NIH_MGC_7
gb_gss:AZ578784           -    96.00  169.65   3.34      377  ! AZ578784 27c09 Shot-gun genomic
gb_est1:AV404426          -    96.00  166.96   4.71      471  ! AV404426 AV404426 Bombyx mori p
gb_gss:BH161414           +    96.00  158.63   13.71     938  ! BH161414 ENTSV42TR Entamoeba hi
gb_est1:AA223267          +    95.00  165.22   5.90      464  ! AA223267 zr08c06.r1 Stratagene
gb_htc:AF314818           +    95.00  148.66   49.30     1826  ! AF314818 Mus musculus FKSG15 (
gb_gss:AQ409086           -    94.50  164.76   6.25      445  ! AQ409086 HS_5059_A2_E04_SP6E RP
gb_est2:BG924582          -    94.50  162.18   8.71      551  ! BG924582 HNC28-1-B11.R HNC (Hum
gb_est2:BM464964          +    94.50  154.12   24.46     1073  ! BM464964 AGENCOURT_6428411 NIH
gb_gss:CNS0003J           +    94.50  153.81   25.46     1101  ! AL050505 Drosophila melanogast
gb_gss:B07740             -    94.00  162.20   8.68      508  ! B07740 8025PI05092795T7 Rhodoba
gb_htc:BC012585           +    94.00  143.89   90.85     2310  ! BC012585 Homo sapiens, clone I
gb_htc:AK013940           +    94.00  141.65   121.06    2780  ! AK013940 Mus musculus 13 d
gb_gss:BH159784           +    93.50  154.11   24.50     916  ! BH159784 ENTRE77TR Entamoeb
gb_gss:CNS06GFN           +    93.50  152.93   28.51     1010  ! AL397641 T3 end of clone A
gb_gss:AQ639443           -    93.00  156.78   17.39     678  ! AQ639443 927P1-12G2.TP 927P
gb_est2:BI087877          +    93.00  153.73   25.73     873  ! BI087877 602852681F1 NIH_MG

seq_name: gb_est1:AI777353

seq_documentation_block:
LOCUS       AI777353               644 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST263761 tomato seed, TAMU Lycopersicon esculentum cDNA clone
            cLEE10E23, mRNA sequence.
ACCESSION   AI777353
VERSION     AI777353.1  GI:5275394
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 644)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
            , Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
            ,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
            ,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato seed tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .644
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEE10E23"
                     /clone_lib="tomato seed, TAMU"
                     /tissue_type="seeds"
                     /dev_stage="quiescent seed"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLEE - Tomato Seed EST Library. Directionally cloned
                     cDNAs inserted into pBlueScript SK(-) at 5' end with
                     EcoRI and 3' end with XhoI site."
BASE COUNT      145 a    159 c    156 g    184 t
ORIGIN


alignment_scores:
           Quality:  167.50              Length:    219
             Ratio:  1.329                 Gaps:      7
 Percent Similarity:  57.534   Percent Identity: 25.114

alignment_block:
 US-09-824-567-2 x AI777353/rev   ..

 Align seq 1/1  to reverse of: AI777353  from: 1  to: 644

     136 AsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSerIleAs 152
         |||      ||||||:::   |||   :::            :::|||||
     627 AACCTGTATAAGAATATTGCTAAAACGGAA.....GCGATCGA 590

     152 pHisPheGlyValHisSerProAsnGluSerThrLeuValValThrLeuG 169
         |                        :::|||:::   :::|||||:
     589 T.....CCGACAACGGTAAAGATTACCCTCA 564

     169 luSerProThrSerHisPheLeuLysLeuLeuAlaLeuProValPhePhe 185
         :::::|||   |||   |||::::::::::||||||   |||:::
     563 AACAGCCGTTCTCAGCGTTTATTAATATTCTTGCCCATCCGGCGACCGCG 514
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 31, 2002, 17:10:54 ; Search time 1882.78 Seconds
(without alignments)
489.047 Million cell updates/sec

Title: US-09-824-567-3
Perfect score: 44
Sequence: 1 ataagaatgcggccgccacc.....gcaagatatcagtgggaatc 44

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	44	100.0	44	6	AX268343	AX268343 Sequence
	2	27.6	62.7	49	6	AX300413	AX300413 Sequence
	3	26.2	59.5	42	6	AX300427	AX300427 Sequence
	4	26	59.1	11944	1	AE001293	AE001293 Chlamydia
	5	25.8	58.6	43	6	AX179663	AX179663 Sequence
	6	25.8	58.6	45	6	AX278234	AX278234 Sequence
	7	25	56.8	45	6	AX300411	AX300411 Sequence
	8	25	56.8	46	6	AX179705	AX179705 Sequence
	9	24.8	56.4	45	6	AX146974	AX146974 Sequence
	10	24.8	56.4	45	6	AX300423	AX300423 Sequence
	11	24.8	56.4	48	6	AX300417	AX300417 Sequence
	12	24.4	55.5	44	6	AX147161	AX147161 Sequence
	13	24.4	55.5	12173	1	AE002315	AE002315 Chlamydia
	14	24	54.5	42	6	AX100501	AX100501 Sequence
	15	24	54.5	42	6	AX268467	AX268467 Sequence
	16	24	54.5	46	6	AX300419	AX300419 Sequence
	17	24	54.5	1235	6	AX300393	AX300393 Sequence
	18	24	54.5	1599	6	AX349501	AX349501 Sequence
	19	24	54.5	1799	6	AX268341	AX268341 Sequence
	20	24	54.5	11648	1	AE001606	AE001606 Chlamydia
C	21	24	54.5	11764	1	AE002216	AE002216 Chlamydop
	22	24	54.5	299650	1	AP002545	AP002545 Chlamydop
	23	23.6	53.6	46	6	AX300425	AX300425 Sequence
	24	23.4	53.2	43	6	AX045141	AX045141 Sequence
	25	23.4	53.2	44	6	AX100530	AX100530 Sequence
	26	23.4	53.2	44	6	AX300415	AX300415 Sequence
	27	23.4	53.2	45	6	AX100526	AX100526 Sequence
	28	23.4	53.2	45	6	AX268509	AX268509 Sequence
	29	23.4	53.2	110000	2	HSY313F4_0	AL023808 Homo sapi
C	30	23.4	53.2	158335	9	AL365199	AL365199 Human DNA
	31	23.4	53.2	200882	2	AL591181	AL591181 Homo sapi
C	32	23.4	53.2	214180	2	AL360082	AL360082 Homo sapi
C	33	23.4	53.2	300000	9	AL645813	AL645813 Human DNA
	34	23.2	52.7	43	6	AX045139	AX045139 Sequence
	35	23.2	52.7	43	6	AX268517	AX268517 Sequence
	36	23	52.3	39	6	AX100509	AX100509 Sequence
	37	23	52.3	42	6	AX068408	AX068408 Sequence
	38	23	52.3	42	6	AX300421	AX300421 Sequence
	39	23	52.3	43	6	AX147650	AX147650 Sequence
	40	23	52.3	43	6	AX300409	AX300409 Sequence
	41	23	52.3	45	6	AX100505	AX100505 Sequence
	42	23	52.3	45	6	AX100513	AX100513 Sequence
	43	23	52.3	45	6	AX179576	AX179576 Sequence
C	44	22.8	51.8	453	6	AX120393	AX120393 Sequence
	45	22.8	51.8	1068	3	PVCLCPRNA	X85127 P.vannamei

ALIGNMENTS

RESULT 1
AX268343
LOCUS AX268343 44 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 3 from Patent WO0174863.
ACCESSION AX268343
VERSION AX268343.1 GI:16541564
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (sites)
AUTHORS Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
TITLE Chlamydia antigens and corresponding dna fragments and uses thereof
JOURNAL Patent: WO 0174863-A 3 11-OCT-2001;
Aventis Pasteur Limited (CA)
FEATURES Location/Qualifiers
source 1. .44
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="5' PCR primer"
BASE COUNT 14 a 11 c 12 g 7 t
ORIGIN

```
  Query Match            100.0%;  Score 44;  DB 6;  Length 44;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-08;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtgggaatc 44
          ||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATAAGAATGCGGCCGCCACCATGCGCAAGATATCAGTGGGAATC 44


RESULT   2
AX300413
LOCUS       AX300413                  49 bp    DNA     linear   PAT 30-NOV-2001
DEFINITION  Sequence 25 from Patent WO0185972.
ACCESSION   AX300413
VERSION     AX300413.1  GI:17381804
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0185972-A 25 15-NOV-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       12 a     14 c     12 g     11 t
ORIGIN


  Query Match            62.7%;  Score 27.6;  DB 6;  Length 49;
  Best Local Similarity  78.6%;  Pred. No. 0.46;
  Matches   33;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagtgggaa 42
          |||||||||||||||||||||||  ||   ||   ||| ||
Db      1 ATAAGAATGCGGCCGCCACCATGCTCACCCTAGGCTTGGAAA 42


RESULT   3
AX300427
LOCUS       AX300427                  42 bp    DNA     linear   PAT 30-NOV-2001
DEFINITION  Sequence 39 from Patent WO0185972.
ACCESSION   AX300427
VERSION     AX300427.1  GI:17381818
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0185972-A 39 15-NOV-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       15 a      9 c     10 g      8 t
ORIGIN


  Query Match            59.5%;  Score 26.2;  DB 6;  Length 42;
  Best Local Similarity  90.3%;  Pred. No. 1.8;
  Matches   28;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagat 31
          |||||||||||||||||||||||||   |||||
Db      1 ATAAGAATGCGGCCGCCACCATGGTAAAGAT 31


RESULT   4
AE001293
LOCUS       AE001293                11944 bp    DNA     linear   BCT 30-OCT-2000
DEFINITION  Chlamydia trachomatis section 20 of 87 of the complete genome.
ACCESSION   AE001293 AE001273
VERSION     AE001293.1  GI:3328597
KEYWORDS    .
SOURCE      Chlamydia trachomatis.
  ORGANISM  Chlamydia trachomatis
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE   1  (bases 1 to 11944)
  AUTHORS   Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,
            Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
            Koonin,E.V. and Davis,R.W.
  TITLE     Genome sequence of an obligate intracellular pathogen of humans:
            Chlamydia trachomatis
  JOURNAL   Science 282 (5389), 754-759 (1998)
  MEDLINE   99000809
   PUBMED   9784136
REFERENCE   2  (bases 1 to 11944)
  AUTHORS   Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
            Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
  TITLE     Comparative genomes of Chlamydia pneumoniae and C. trachomatis
  JOURNAL   Nat. Genet. 21 (4), 385-389 (1999)
  MEDLINE   99206606
   PUBMED   10192388
REFERENCE   3  (bases 1 to 11944)
  AUTHORS   Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,
            Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
            Koonin,E.V. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-1998) Program in Infectious Diseases, University
            of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES             Location/Qualifiers
     source          1. .11944
                     /organism="Chlamydia trachomatis"
                     /strain="D/UW-3/CX"
                     /db_xref="taxon:813"
     gene            complement(105. .878)
                     /gene="CT192"
     CDS             complement(105. .878)
                     /gene="CT192"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAC67784.1"
                     /db_xref="GI:3328598"
                     /translation="MFFSGSVETLPPSTCLISKKTRKEEFVRSVGQEASQNTLSWTIR
                     PRIPSIIQGSKEVSLALFVLGTVLAIVGACAAAVGGAFSVCLGVLFLGGVVLATGLLL
                     AVLEFCHIRSSMEKYVILTKQDLFKEPVIQEEQATPLIEEASYTCEPGIPLSGPEEVQ
                     QERPVILQKDLDLSHAPKYIAVGSHVVELVKAGKIGRNGERLLEEGIDTDQNFVRVWD
                     ELFILGQRGEMVRLDGFCCKVLPKKSKSESINNLVSNDC"
     gene            complement(905. .2023)
                     /gene="tgt"
                     /note="CT193"
     CDS             complement(905. .2023)
                     /gene="tgt"
                     /codon_start=1
                     /transl_table=11
                     /product="Queuine tRNA Ribosyl Transferase"
                     /protein_id="AAC67785.1"
                     /db_xref="GI:3328599"
                     /translation="MALRFEILHQSKKSRARVGRIETAHGYIDTPAFVPVATNGALKG
                     VLDHSNIPLMFCNTYHLIVHPGAEAIAAMGGLHQFIGRNAPIITDSGGFQIFSLAYGS
                     VAEEIKSCGKKKGGNTIIKVNDDGVHFKSYRDGRKLFLSPEISVQAQKDLGADIILPL
                     DELLPFHADPTYFHQSSQRTYVWEKRSLDYHLKNPGIQSMYGVIHGGTFPDQRKLGCK
                     FVEDLPFDGSAIGGSLGKNLQDIVEVVGVTAANLSAERPRHLLGIGDLPSIWATVGFG
                     IDSFDSSYPTKAARHGMILTSQGPLKINNQRYSSDLNPIEPGCSCLACSQGITRAYLR
                     HLFKVHEPNAGIWASIHNMHHMQKVMREIREGILNDRI"
     gene            complement(2150. .3562)
```

/gene="mgtE"
/note="CT194"
CDS complement(2150. .3562)
/gene="mgtE"
/codon_start=1
/transl_table=11
/product="Mg++ Transporter (CBS Domain)"
/protein_id="AAC67786.1"
/db_xref="GI:3328600"
/translation="MDSKTSHLDDELCFKLEEAFDTLTAGEHSQDLTSIVSVYNPIDL
AYAVSCLPSDSRSILYKNLDSIASKIAFIINTDSASRWAIFRNLSDGEICALIEQMPP
DEAIWVLDDIPDRRYRRILDLIDVKKALKIRDLQKHGRNTAGRLMTNEFFAFLMETTV
KEVATCIRNNPGIDLTRLVFVLDFKGELQGFVTDRSLIIASPEMPLKQIMRPIEHKVL
ADTTREEVVDLVERYKVAVLPVVDEENFLIGAITYEDVVETIEDIADETIARMAGTTE
DVGYHDCHVVQRFLLRAPWLLITLCAGLVSASVMAYFQKIAPTLLAMVIFFIPLVNGL
SGNVGVQCSTILVRSMATGTLSFGRRRETILKEMSIGLLTGVALGILCGLVVCCMGCL
GLGLFATGGVQLGVTVSVGILGASLTATTLGVLSPFFFAKIGVDPALASGPIVTALND
IMSMVIFLLITGTLNVLFFK"
gene complement(4001. .5092)
/gene="CT195"
CDS complement(4001. .5092)
/gene="CT195"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67787.1"
/db_xref="GI:3328601"
/translation="MVSMSLNLPPAEVRLRPVTAKSFPNSSAWRATQKRITGCSLNAS
SYRPSSARVISFVVGVLVVLGWICFRIAYLANSRVLTFPKMFAIAILPFILFGGGLAI
LFRIAGKVDVLYGKKIQPFISRQWERVILCEKEGEFIRPIQEPDMYMDVSLLDRTGSG
IAPVYTYPPMDARTVICVVTSILLPIFSIVRMLYNIFRFFIVPFYILFQMVRQNYQTD
IPKEERFVCSDIVREMTRSLLQAVKAPFYGAVCYLANLYGLLNPLSGRVVLASLERDW
NNDVIRSRGVWGIFCEKNYMFEGGGTRSGLGQNAWYLLGCFQPVQLFLLKDGVIISGA
RPSIQSFPESKEYLASFLYGAVPGRLAGF"
gene 5316. .5636
/gene="CT196"
CDS 5316. .5636
/gene="CT196"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67788.1"
/db_xref="GI:3328602"
/translation="MATTVNPNYSLPFCEKMVSSATLAAHSFFNHIKLLIPLLVGYFC
ILLGALILTGVITTIPSIAASYFLSLGVTLVIVGAGLCAAFKRPLFSVTQSKASTLLH
ISCD"
gene 5712. .6728
/gene="gcp_1"
/note="CT197"
CDS 5712. .6728
/gene="gcp_1"
/codon_start=1
/transl_table=11
/product="O-Sialoglycoprotein Endopeptidase"
/protein_id="AAC67789.1"
/db_xref="GI:3328603"
/translation="MLTLGLESSCDETSCSLVQNGKILANKIASQDIHASYGGVIPEL
ASRAHLQTFPELLTAATQSAGVSLEDIELISVANTPGLIGALSIGVNFAKGLASGLKR
PLIGVNHVEAHLYAACMEAPATQFPALGLAISGAHTSLFLMPDATTFLLIGKTRDDAI
GETFDKVARFLGLPYPGGQKLEELAREGDADAFAFSPARVSGYDFSFSGLKTAVLYAL
KGNNSSAKAPFPEVSETQKRNIAASFQKAVFMTIAQKLPDIVKTFSCESLIVGGGVAN
NSYFRRLLNQICSLPIYFPSSQLCSDNAAMIAGLGERLFCNRTHVSKEVIPCARYQWE
SACS"
gene 6692. .8248
/gene="oppA_3"
/note="CT198"
CDS 6692. .8248
/gene="oppA_3"
/codon_start=1
/transl_table=11
/product="Oligopeptide Binding Protein"
/protein_id="AAC67790.1"
/db_xref="GI:3328604"
/translation="MRKISVGICLLLALATSGCSKSSSNATHRSPATHTVAVSVKDDP
RTFDPREVRLLSDINLIHHLYEGLVQETPSGEVFPALAESFFLSEDKKTYTFNLKKAF
WSNGDLITAHDFVRSWNDVLQNRVASIYSFAFLPIDVNKDSGFFAKDDHTLVINLLTP
TPHFLKLLTLPVFYPVHSQHQIRKEEKSLPISTGAFFLKEKKDRRWLKLEKSPYYYNK
DQVAVQEICIHIIPDQQTASALFNQGKLDWQGLPWGHSIPQETLATTNKRRAPRSFDI
SGTSWLTFNTAKKPFSHSKLRQALSLVLNKEALASLAFVKPAKHLLPAHLHTYPEQPS
YKQQEAITLAKSLLEEALTELNMTIEDLEKYPLTFSATSTMNSQIAQMLRDQWRRSLG
ITFPICGKEYALLQNDLIGNTFFMSIGGWFADFSDPLAFLSIFSSKGVKPYALQDPQF
DQLILSIETEKNPQKRSALISEASLYIERQNVIEPLYHDVFHYTTNNKLSFVRLHPSG
LVDMRYAKNS"
gene 8407. .9348
/gene="oppB_1"
/note="CT199"
CDS 8407. .9348
/gene="oppB_1"
/codon_start=1
/transl_table=11
/product="Oligopeptide Permease"
/protein_id="AAC67791.1"
/db_xref="GI:3328605"
/translation="MLSYIKRRLLFNLLSLWVVVTLTFFIIKTIPGDPFNDENGNILS
SETLALLKNRYGLDKPLFTQYLIYLKCLLTLDFGESLIYKDRTVISIIAAALPSSAIL
GLESLCLSLFGGITLGILAAFYKKSCGRTIFFSSVIQISVPAFVIGAFLQYVFAIKYS
CLPIACWGNFSHTLLPSIALAITPMAFITQLTCASVSANLKKDYVLLAYAKGLSPFKV
LIKHILPYALFPVISYSAFLITTLMTGTFSIENLFCIPGLGKWFICSIKQRDYPITLG
LSVFYGAFFMLTSLCCDLLQAWIDPQIRYSYGKERSK"
gene 9380. .10225
/gene="oppC_1"
/note="CT200"
CDS 9380. .10225
/gene="oppC_1"
/codon_start=1
/transl_table=11
/product="Oligopeptide Permease"
/protein_id="AAC67792.1"
/db_xref="GI:3328606"
/translation="MFRSSSSHTWRYIRTNKMLVLGCLTLAVLVLAATFLPYLYPDYE
RTFPEHALQSPSKAFPFGTDNLGRCMLARTLQGIRLSLLIAVSATIIDVCLGLLWSTL
ALASGKKVADIMSRITEILFSIPRIPVIILLLVIFNHGILPLILAMTITGWIPIARII
YGQPLLLENKEFVLSARALQASTFHILRKHLLPNSLGPIISTLIFTIPNAVYTEAFIS
FLGLGIQPPYASLGTLVKEGIHSLAYHPWLFFIPSFFMIIVSVSFNCIGEGLRTKLLE
ENTLV"
gene 10218. .11051
/gene="oppD"
/note="CT201"
CDS 10218. .11051
/gene="oppD"
/codon_start=1
/transl_table=11
/product="Oligopeptide Transport ATPase"
/protein_id="AAC67793.1"
/db_xref="GI:3328607"
/translation="MSEDLLKIDNLVVSVKDSNQRLVNHLSLTIKRCQSMALVGENGS
GKTTVSKAVLGFLPDNCYIQSGRILYSSTDITRLSRRQLQTIRGKKIATIFQNAMGTL
TPSMRVGAQIVETLRHHFDMSKEEAFSKARELLESVHIESPDRCLQLYPFELSGGMCQ
RVSIAIALATNPELIIADEPSTALDSISQAQVLRVLTQIHQNHSTALLLITHNLALVS
ELCEEMAIIRYGEIVEQGPVQELLHSPSHPYTQQLIRAIPKIPSPSYLSPKEPLATTA
Y"
gene 11066. .11809

Query Match 59.1%; Score 26; DB 1; Length 11944;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 ccatgcgcaagatatcagtgggaatc 44
||||||||||||||||||||||||||
Db 6690 CCATGCGCAAGATATCAGTGGGAATC 6715

RESULT 5
AX179663
LOCUS AX179663 43 bp DNA linear PAT 06-AUG-2001
DEFINITION Sequence 3 from Patent WO0146225.

```
ACCESSION   AX179663
VERSION     AX179663.1  GI:15132084
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 43)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0146225-A 3 28-JUN-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .43
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       16 a      9 c      6 g     12 t
ORIGIN


  Query Match           58.6%;  Score 25.8;  DB 6;  Length 43;
  Best Local Similarity  81.1%;  Pred. No. 2.7;
  Matches   30;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcagt 37
          ||||||||||||||||||||||||  ||  ||| | |
Db      1 ATAAGAATGCGGCCGCCACCATGAAAAAATTATTATT 37


RESULT   6
AX278234
LOCUS       AX278234                 45 bp    DNA     linear   PAT 01-NOV-2001
DEFINITION  Sequence 3 from Patent WO0175114.
ACCESSION   AX278234
VERSION     AX278234.1  GI:16605283
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0175114-A 3 11-OCT-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       12 a      14 c      8 g     11 t
ORIGIN


  Query Match           58.6%;  Score 25.8;  DB 6;  Length 45;
  Best Local Similarity  93.1%;  Pred. No. 2.7;
  Matches   27;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaag 29
          |||||||||||||||||||||||||  |||
Db      1 ATAAGAATGCGGCCGCCACCATGCAGAAG 29


RESULT   7
AX300411
LOCUS       AX300411                 45 bp    DNA     linear   PAT 30-NOV-2001
DEFINITION  Sequence 23 from Patent WO0185972.
ACCESSION   AX300411
VERSION     AX300411.1  GI:17381802
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0185972-A 23 15-NOV-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       13 a      11 c      13 g      8 t
ORIGIN


  Query Match           56.8%;  Score 25;  DB 6;  Length 45;
  Best Local Similarity  100.0%;  Pred. No. 5.9;
  Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcg 25
          |||||||||||||||||||||||||
Db      1 ATAAGAATGCGGCCGCCACCATGCG 25


RESULT   8
AX179705
LOCUS       AX179705                 46 bp    DNA     linear   PAT 06-AUG-2001
DEFINITION  Sequence 3 from Patent WO0146226.
ACCESSION   AX179705
VERSION     AX179705.1  GI:15132097
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 46)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0146226-A 3 28-JUN-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .46
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       21 a      13 c      6 g      6 t
ORIGIN


  Query Match           56.8%;  Score 25;  DB 6;  Length 46;
  Best Local Similarity  84.8%;  Pred. No. 5.9;
  Matches   28;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatat 33
          |||||||||||||||||||||||||  |  || |
Db      1 ATAAGAATGCGGCCGCCACCATGCAAACAATCT 33


RESULT   9
AX146974
LOCUS       AX146974                 45 bp    DNA     linear   PAT 08-JUN-2001
DEFINITION  Sequence 3 from Patent WO0136455.
ACCESSION   AX146974
VERSION     AX146974.1  GI:14346245
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 45)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0136455-A 3 25-MAY-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
```

```
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       16 a     11 c     10 g      8 t
ORIGIN


  Query Match             56.4%;  Score 24.8;  DB 6;  Length 45;
  Best Local Similarity   80.6%;  Pred. No. 7.2;
  Matches   29;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaagatatcag 36
          |||||||||||||||||||||   |   | ||||
Db      1 ATAAGAATGCGGCCGCCACCATGGTAACAGTTTCAG 36


RESULT  10
AX300423
LOCUS       AX300423                  45 bp    DNA     linear   PAT 30-NOV-2001
DEFINITION  Sequence 35 from Patent WO0185972.
ACCESSION   AX300423
VERSION     AX300423.1  GI:17381814
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0185972-A 35 15-NOV-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .45
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       13 a     14 c     10 g      8 t
ORIGIN


  Query Match             56.4%;  Score 24.8;  DB 6;  Length 45;
  Best Local Similarity   92.9%;  Pred. No. 7.2;
  Matches   26;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaa 28
          ||||||||||||||||||||||||| || |
Db      1 ATAAGAATGCGGCCGCCACCATGAGCTA 28


RESULT  11
AX300417
LOCUS       AX300417                  48 bp    DNA     linear   PAT 30-NOV-2001
DEFINITION  Sequence 29 from Patent WO0185972.
ACCESSION   AX300417
VERSION     AX300417.1  GI:17381808
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0185972-A 29 15-NOV-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .48
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       17 a     11 c      8 g     12 t
ORIGIN


  Query Match             56.4%;  Score 24.8;  DB 6;  Length 48;
  Best Local Similarity   92.9%;  Pred. No. 7.3;
  Matches   26;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgcaa 28
          ||||||||||||||||||||||||| || |
Db      1 ATAAGAATGCGGCCGCCACCATGAGAAA 28


RESULT  12
AX147161
LOCUS       AX147161                  44 bp    DNA     linear   PAT 08-JUN-2001
DEFINITION  Sequence 3 from Patent WO0136457.
ACCESSION   AX147161
VERSION     AX147161.1  GI:14346332
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 44)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0136457-A 3 25-MAY-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .44
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       15 a     17 c      7 g      5 t
ORIGIN


  Query Match             55.5%;  Score 24.4;  DB 6;  Length 44;
  Best Local Similarity   96.2%;  Pred. No. 11;
  Matches   25;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 ataagaatgcggccgccaccatgcgc 26
          |||||||||||||||||||||||||
Db      1 ATAAGAATGCGGCCGCCACCATGCAC 26


RESULT  13
AE002315
LOCUS       AE002315              12173 bp    DNA     linear   BCT 26-MAY-2000
DEFINITION  Chlamydia muridarum, section 46 of 85 of the complete genome.
ACCESSION   AE002315 AE002160
VERSION     AE002315.2  GI:8163226
KEYWORDS    .
SOURCE      Chlamydia muridarum.
  ORGANISM  Chlamydia muridarum
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE   1  (bases 1 to 12173)
  AUTHORS   Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
            White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
            Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
            Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
            McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
  TITLE     Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
            pneumoniae AR39
  JOURNAL   Nucleic Acids Res. 28 (6), 1397-1406 (2000)
  MEDLINE   20150255
   PUBMED   10684935
REFERENCE   2  (bases 1 to 12173)
  AUTHORS   Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
            White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
            Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
            Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
            McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Jun 1, 2000 this sequence version replaced gi:7190506.
FEATURES             Location/Qualifiers
```

```
source          1. .12173
                /organism="Chlamydia muridarum"
                /db_xref="taxon:83560"
                /note="synonym: Chlamydia trachomatis MoPn"
gene            complement(97. .792)
                /gene="TC0464"
CDS             complement(97. .792)
                /gene="TC0464"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF39314.1"
                /db_xref="GI:7190507"
                /translation="MDSSCALKNLSSCEVSYRIPEVIATRKEVALVLIVLGTVLAAIG
                ACAVAVISGVWSIGLGTLVVGGVLLATGFLFALMEICRAQYGRRQQKIQTQQEVLESS
                PIREEATLPFIEQAVYSHEPELPLSSQEEVRCHRPIILQRGSKPAETPKFIAVGNSVV
                ELVKVGMIGPMGRRGNVNPGQTLVRLWDELFALGRMGELVRLDGFCCKVLPATLGDTS
                VSKTSADRASSED"
gene            complement(1227. .2345)
                /gene="TC0465"
CDS             complement(1227. .2345)
                /gene="TC0465"
                /note="similar to GB:AL009126; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="queuine tRNA-ribosyltransferase"
                /protein_id="AAF39315.1"
                /db_xref="GI:7190508"
                /translation="MALRFEILHQSKKSRARVGRIETEHGCIDTPAFVPVATNGALKG
                VLDHSNIPLMFCNTYHLIVHPGAEAVAAMGGLHQFIGRNAPIITDSGGFQIFSLAYGS
                VAEEIKSCGKKKGENSIIKINDEGVWFKSYRDGRKLFLSPEVSVQAQKHLGADIIIPL
                DELLPFHTDPAYFQQSSQRTYAWEKRSLDYHLANPGYQSMYGVIHGGTFPDQRKLGCQ
                FVEDLPFDGSAIGGSLGKNLRDIVGVVDVTTANLSIERPRHLLGIGDLPSIWATVGFG
                IDSFDSSYPTKAARHGMILTSQGSLKINNQRYASDLNPIEPGCACPACSQGISRAYLR
                HLFKVHEPNAGIWASIHNMHYMQKIMSKIREKILNDQL"
gene            complement(2467. .3939)
                /gene="TC0466"
CDS             complement(2467. .3939)
                /gene="TC0466"
                /note="magnesium transporter; identified by match to TIGR
                protein family HMM TIGR00400"
                /codon_start=1
                /transl_table=11
                /product="magnesium transporter"
                /protein_id="AAF73559.1"
                /db_xref="GI:8163227"
                /translation="MGEQNRNEEKLDVALTSGNLMDSKTSHLDDELSFKLEEAFDTLT
                AGEHSQDLTSIVSAYNPIDLAYAVSCLPSDSRSILYKSLDSIASKIAFIINTDSASRW
                AIFRNLSDGEICALIEQMPPDEAIWVLDDIPDRRYRRILDLIDVKKALKIRDLQKHGR
                NTAGRLMTNEFFAFLMETTVKEVATCIRNNPGVDLTRLVFVLDFKGELQGFVTDRSLI
                IASPEMPLKQIMRPVEHKVLADTTREEVVDLVERYKVAVLPVVDEENFLIGAITYEDV
                VETIEDIADETIARMAGTTEDVGYHDCHVVQRFLLRAPWLLITLCAGLVSASVMAYFQ
                KIAPTLLAMVIFFIPLVNGLSGNVGVQCSTILVRSMATGTLSFGRRRETILKEMSIGL
                LTGVALGILCGLVVCCMGCLGLGLFSTGGVQLGVTVSVGILGASLTATTLGVLSPFFF
                AKIGVDPALASGPIVTALNDIMSMVIFLLITGTLNVLFFK"
gene            4208. .4366
                /gene="TC0467"
CDS             4208. .4366
                /gene="TC0467"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF39316.1"
                /db_xref="GI:7190509"
                /translation="MDKAHTVHFPPQLQLYDEDFTVDLQAKSLGLLKTVAVYAKLAAS
                QPSQGSPI"
gene            complement(4328. .5410)
                /gene="TC0468"
CDS             complement(4328. .5410)
                /gene="TC0468"
                /note="conserved hypothetical protein; identified by
                Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AAF39317.1"
                /db_xref="GI:7190510"
                /translation="MFLNLPSAEVRLRPVTTESFPRASSWKATQKRITSCTLTNPFYK
                TSLTRIVCFVVGILIVLGAICCRIAYLANAGLLFSPKILAIALSSLVLLGGGIGVLCK
                IARKIDVLYGEKIQPFAIRKWEQVILCEKRGQTIRPIQDPDMYMDVSLLDKQGSGIAP
                VYTYPPAKARTAICFIACILLPIITVVRVLYNAFRFLLIPFYIVFQMIRQLYQEDLPF
                EEQFICSDIFREMSRSFVQAVKAPFYGVACYLASLYGLLNPLSGRVIMASVERDWNND
                VIRSRGIWGIFCEKNCLLEGGGTRSGLGQHAWYLLGCFQPFRLFLLKDGEIVSGARPS
                IQAFPENKEYLTSYLYGAALGRLTSR"
gene            5581. .5946
                /gene="TC0469"
CDS             5581. .5946
                /gene="TC0469"
                /note="hypothetical protein; identified by Glimmer2;
                putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF39318.1"
                /db_xref="GI:7190511"
                /translation="MNYLFSLNNQKLKARMQLTIVNTNSRIFPFSNERILPGSTLSTD
                STFSYIKILIPLLAGSICLLLGALILAGTITTVPAIAASYFLSLGVTLVVAGIGLCSV
                FRPHLFSINQSQDLHIIYY"
gene            6012. .7034
                /gene="TC0470"
CDS             6012. .7034
                /gene="TC0470"
                /note="O-sialoglycoprotein endopeptidase; identified by
                match to PFAM protein family HMM PF00814"
                /codon_start=1
                /transl_table=11
                /product="O-sialoglycoprotein endopeptidase"
                /protein_id="AAF73560.1"
                /db_xref="GI:8163228"
                /translation="MLTLGLESSCDETSCALVENGKILANRIASQDIHAAYGGVIPEL
                ASRAHLQIFPKLLAAVAQDAEVSLEDVELISVANTPGLIGALSVGVNFAKGLASGLKK
                TLIGVNHVEAHLYAACLEEPSIRFPALGLAISGAHTSLFLMPNATTFLLIGKTRDDAI
                GETFDKVARFLGLPYPGGQKLEELAQDGDEEAYPFSRAKVSGNDFSFSGLKTAVLYAL
                KGNNSSAKAPFPEVSETQKRNIAASFQKAAFMTIAQKLPDIVKAFSCESLIVGGGVAN
                NRYFRRLLNQTCSLPTYFPSSQLCSDNAAMIAGLGERLFCNQTYVSKEVIPCARYQWE
                SACLSH"
gene            6992. .8554
                /gene="TC0471"
CDS             6992. .8554
                /gene="TC0471"
                /note="similar to GB:M57689 SP:P24141 GB:X56347 PID:143603
                PID:40005; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="peptide ABC transporter, periplasmic
                peptide-binding protein, putative"
                /protein_id="AAF39319.1"
                /db_xref="GI:7190512"
                /translation="MRKVSVGICLLIALATAITGCSKSSSNKSNHSSSNQSVSVSMKD
                DPRTFDPREVRLLSDINLIHHLYEGLVQETPSGEVFPALAESFFLSEDKKTYTFHLKK
                ALWSNGDLITAHDFVRSWNDVLQNRIASIYSFAFLPIDLSKDCGFFAKDNHTLVINLH
                TPTPHFLKLLTLPVFYPVHPEHQIRNEAKALPISTGAFCLKEKKDRRWLKLEKNPYYY
                NKEQVAIQEIHIHVIPDQQTASALFKQGKLDWQGLPWGHSIPQEALATANKRRTPQSF
                DISGTSWLTFNTSKVPFSHPKLRQALSLVLNKEALASPTFVKPAKHLLPTHLHTYPEQ
                PTYKQQEAVILAKTLLQEALTDLNMTIKDLEKCPLIFSATSSVNSQMAQMMRDQWRRI
                LGITFPICGKEYALLQNDLTTNTFFMSINGWFADFSDPLAFLSVFSSKGIKPYALQDP
                LFDQLILSIETEKDPRKRISLISEASLYLEKQNIIEPLYHDVFHYAANNKLSFVRLHP
                SGLVDMRYAKNS"
gene            8668. .9612
                /gene="TC0472"
CDS             8668. .9612
```

```
                     /gene="TC0472"
                     /note="similar to SP:P24138 GB:X56347 PID:580897
                     GB:AL009126; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="peptide ABC transporter, permease protein"
                     /protein_id="AAF39320.1"
                     /db_xref="GI:7190513"
                     /translation="MLGYIKKRLLFNLLSLWVVVTLTFFIIRTIPGDPFNDENGNVLS
                     EEILAILNKRYGLDKPLFTQYLIYLKCLLTLDFGESLIYKDRTVIGIITTALPSSAIL
                     GLESLCLALFGGITLGILAAFYKRGCGRTIFFSSIIQISVPAFVIGAFLQYIFAIKYS
                     LLPIACWGDFSHTLLPSIALAIAPMAFITQLTYASVSASLKKDYVLLAYAKGLSPLKV
                     LIKHVLPYALFPVISYSAFLITTLMTGTFSIENLFCIPGLGKWFICSIKQRDYPMTLG
                     LSVFYGAFFMLISLLCDLLQAWIDPQIRYSYGKEHSKQ"
     gene            9618. .10481
                     /gene="TC0473"
     CDS             9618. .10481
                     /gene="TC0473"
                     /note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606
                     PID:40007; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="peptide ABC transporter, permease protein"
                     /protein_id="AAF39321.1"

  Query Match             55.5%;  Score 24.4;  DB 1;  Length 12173;
  Best Local Similarity   96.2%;  Pred. No. 20;
  Matches   25;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      19 ccatgcgcaagatatcagtgggaatc 44
           |||||||||| ||||||||||||||
Db    6990 CCATGCGCAAGGTATCAGTGGGAATC 7015


RESULT  14
AX100501
LOCUS       AX100501                 42 bp    DNA     linear   PAT 10-APR-2001
DEFINITION  Sequence 3 from Patent WO0121803.
ACCESSION   AX100501
VERSION     AX100501.1  GI:13619514
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 42)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     I chlamydia /i antigens and corresponding dna fragments and uses
            thereof
  JOURNAL   Patent: WO 0121803-A 3 29-MAR-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       15 a     11 c     10 g      6 t
ORIGIN


  Query Match             54.5%;  Score 24;  DB 6;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgc 24
           |||||||||||||||||||||||
Db       1 ATAAGAATGCGGCCGCCACCATGC 24


RESULT  15
AX268467
LOCUS       AX268467                 42 bp    DNA     linear   PAT 29-OCT-2001
DEFINITION  Sequence 3 from Patent WO0175113.
ACCESSION   AX268467
VERSION     AX268467.1  GI:16541650
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
  TITLE     Chlamydia antigens and corresponding dna fragments and uses thereof
  JOURNAL   Patent: WO 0175113-A 3 11-OCT-2001;
            Aventis Pasteur Limited (CA)
FEATURES             Location/Qualifiers
     source          1. .42
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="5' PCR primer"
BASE COUNT       13 a     13 c      9 g      7 t
ORIGIN


  Query Match             54.5%;  Score 24;  DB 6;  Length 42;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ataagaatgcggccgccaccatgc 24
           |||||||||||||||||||||||
Db       1 ATAAGAATGCGGCCGCCACCATGC 24


Search completed: July 31, 2002, 19:31:22
Job time: 8428 sec
```